

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 1, 2004, 23:04:53 ; Search time 4244.29 Seconds  
(without alignments)  
10279.384 Million cell updates/sec

Title: US-10-668-047-1  
Perfect score: 1461  
Sequence: 1 gtcatccgacgacgacac.....ttccggaatcgctctgtga 1461

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estmu:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_hic:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hic:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estum:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_mam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rtd:\*
  - 26: em\_gss\_phg:\*
  - 27: em\_gss\_vrl:\*
  - 28: gb\_gss1:\*
  - 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	47.4	3.2	1546	28	BZ568822
2	45.8	3.1	794	28	BZ560168
3	44.8	3.1	922	9	AL571419
4	43.6	3.0	925	29	CNS0091P

C 5	42.6	2.9	690	13	CA136496
C 6	41.6	2.8	718	13	CA131027
C 7	41.4	2.8	1101	29	CNS017SY
8	41.4	2.8	1201	13	EX381961
9	41.2	2.8	508	12	BM727423
C 10	41.2	2.8	770	28	BZ563177
C 11	41.2	2.8	1347	28	BZ574354
C 12	41	2.8	579	12	BM712537
C 13	41	2.8	596	12	BJ469381
C 14	41	2.8	661	12	BJ472412
C 15	40.8	2.8	1201	13	EX356664
C 16	40.2	2.8	721	14	CD494071
C 17	40.2	2.8	1176	14	CD502787
C 18	39.8	2.7	802	12	BI183587
C 19	39.8	2.7	882	29	CG369642
20	39.8	2.7	885	13	EX425603
21	39.8	2.7	958	13	CA151619
22	39.6	2.7	514	10	BF594332
23	39.6	2.7	901	29	CG353488
24	39.6	2.7	921	29	CG446568
25	39.4	2.7	459	12	BM696019
26	39.4	2.7	1029	9	AL528764
C 27	39.2	2.7	625	29	CC711484
C 28	39.2	2.7	806	28	BZ402559
C 29	39.2	2.7	809	29	CC627722
C 30	39.2	2.7	918	29	CG445606
C 31	39.2	2.7	971	29	CG240394
C 32	39.2	2.7	1198	12	BI957046
C 33	39	2.7	949	29	CC930329
C 34	38.8	2.7	391	14	CB879070
C 35	38.8	2.7	490	14	CD036955
C 36	38.8	2.7	507	12	BM075546
C 37	38.8	2.7	524	14	CD036685
C 38	38.8	2.7	534	12	BM335322
C 39	38.8	2.7	542	12	BM335465
C 40	38.8	2.7	572	12	BM335137
C 41	38.8	2.7	574	12	BJ088126
C 42	38.8	2.7	657	12	BM349453
C 43	38.8	2.7	729	28	BZ528408
C 44	38.8	2.7	787	29	CG360243
C 45	38.8	2.7	837	10	BF621278

ALIGNMENTS

RESULT 1  
BZ568822  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BZ568822 1546 bp DNA linear GSS 17-DEC-2002  
pacs2-164\_8106.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
pacs2-164\_8106, genomic survey sequence.  
BZ568822  
BZ568822.1 GI:27202436  
GSS.  
Pseudomonas aeruginosa  
Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
1 (bases 1 to 1546)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
J. Bacteriol. (2002) In press  
Contact: Chris K. Raymond  
Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES

Location/Qualifiers

```
source
1. .1546
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pac82-164 8106"
/clone_lib="pac82-164"
/notes="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN
Query Match 3.2%; Score 47.4; DB 28; Length 1546;
Best Local Similarity 53.2%; Pred. No. 0.45; Indels 2; Gaps 1;
Matches 123; Conservative 0; Mismatches 106;

Qy 149 GAACGGTGGCGGTGCTGCACCCAGAGCTGGTGTGGGTCTCTGACCGTCCCTCTGTCG 208
Db 79 GAGCTGATCGAGATCTGCTTGCAGGAGAGGCTGCGGTCGCAAGACGCTGGTCTATTTCG 138

Qy 209 ATCGGCTGCTACTGAACGCCCGGCTGCGGTCCAGCTAGTTGCGGTGGAAACCGACC 268
Db 139 GTCTACACCGGCACCGCGCACCCACGTCGCGCTTGAAGGTGCTGC--TGGAGCAGGAAG 196

Qy 269 CCTTTGCTGCTTACCTGCGGCCACCCCTGGAGGAATGTGGAAGCTTTACGGCATCT 328
Db 197 GCTTCAAGGTGGCGGTCTGCGCGCAAGCTGGATGCTCCCGCGGAGGACTGGATCG 256

Qy 329 CCTACGACCTAGTCGAGCGGCACTATTGTGTTAAACCAAGGGGCCAAGCTGG 379
Db 257 CCGAGCAGTTGGACCGCGGCATCGACGTGTGATCAGCAACCCCGAGCTGG 307

RESULT 3
AL5711419 922 bp mRNA linear EST 31-MAY-2003
LOCUS
DEFINITION
AL5711419 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI023YB23 3-PRIME, mRNA sequence.
AL5711419.2 GI:31292819
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 922)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
On Feb 16, 2001 this sequence version replaced gi:12928696.
Contact: Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI023JAL2NP1.
Location/Qualifiers
1. .922
FEATURES
Source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI023YB23"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 3.1%; Score 44.8; DB 9; Length 922;
Best Local Similarity 28.0%; Pred. No. 1.7;
Matches 60; Conservative 66; Mismatches 88; Indels 0; Gaps 0;

Qy 500 GGGTGCAGCAGTCATTTTCGCTCAAGAGAGCGGGGGGGGTTTCATGTTCTTCGAT 559
Db 611 GCGMARGTMRVGMVGKGMGMAGVTTMRMGGMGGTGGTGGTGGTGGTGGTGGTGGTGGT 670

Qy 560 CTGTGGCGAACGGCCCTTACTATCTGTCATTTTCGCTGATGATGACCGCGGTAGTC 619
Db 671 GTMMRGTMAMRGKGTGSRMMTTTGMVHTGMAGGGTGTGTGTGTGTGTGTGTGTGT 730

Qy 620 TCGATATACCTTCATGTGTTTCGAAAGTAGAACCAAGTATTTTCGGGACACGAGGTAAGC 679
Db 731 RTGGMATMTWTTTWTGGTTTGTGMRMRMATGMMAGTGGTTMMMMMMMMTGGVGTGGM 790
```

```
source
1. .1546
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pac82-164 8106"
/clone_lib="pac82-164"
/notes="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN
Query Match 3.2%; Score 47.4; DB 28; Length 1546;
Best Local Similarity 53.2%; Pred. No. 0.45; Indels 2; Gaps 1;
Matches 123; Conservative 0; Mismatches 106;

Qy 149 GAACGGTGGCGGTGCTGCACCCAGAGCTGGTGTGGGTCTCTGACCGTCCCTCTGTCG 208
Db 389 GAGCTGATCGAGATCTGCAAGCAGGAGAGGCTGCGGTCGCAAGACGCTGGTCTATTTCG 448

Qy 209 ATCGGCTGCTACTGAACGCCCGGCTGCGGTCCAGCTAGTTGCGGTGGAAACCGACC 268
Db 449 GTCTACACCGGCACCGCGCACCCACGTCGCGCTTGAAGGTGCTGC--TGGAGCAGGAAG 506

Qy 269 CCTTTGCTGCTTACCTGCGGCCACCCCTGGAGGAATGTGGAAGCTTTACGGCATCT 328
Db 507 GCTTCAAGGTGGCGGTCTGCGCGCAAGCTGGATGCTCCCGCGGAGGACTGGATCG 566

Qy 329 CCTACGACCTAGTCGAGCGGCACTATTGTGTTAAACCAAGGGGCCAAGCTGG 379
Db 567 CCGAGCAGTTGGACCGCGGCATCGACGTGTGATCAGCAACCCCGAGCTGG 617

RESULT 2
BZ560168 794 bp DNA linear GSS 17-DEC-2002
LOCUS
DEFINITION
pac82-164_2210.x1 pac82-164 Pseudomonas aeruginosa genomic clone
pac82-164_2210, genomic survey sequence.
BZ560168
ACCESSION
VERSION
KEYWORDS
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1. (bases 1 to 794)
AUTHORS
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol. (2002) In press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .794
FEATURES
Source
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pac82-164_2210"
/clone_lib="pac82-164"
/notes="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN
Query Match 3.1%; Score 45.8; DB 28; Length 794;
Best Local Similarity 52.8%; Pred. No. 0.86;
Matches 122; Conservative 0; Mismatches 107; Indels 2; Gaps 1;
```

[illegible]

QY	328	TCCTACGACCATTGTGCAGCGCACTATTTCGTTTACCAAGGGGCCAAGCTGATGG	383
Db	870	CSSSVSSSKSSASSVSSSGSVSSNSSSASKSSSGSVSSGSGSGSVSG	925
RESULT 5			
LOCUS	CA136496/c		
DEFINITION	SCQGR1043C05.9 RT1 Saccharum officinarum cDNA clone SCQGR1043C05		
ACCESSION	5', mRNA sequence.		
VERSION	CA136496		
KEYWORDS	CA136496.1 GI:35025554		
SOURCE	EST.		
ORGANISM	Saccharum officinarum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD		
TITLE	clade; Panicoideae; Andropogoneae; Saccharum.		
JOURNAL	1 (bases 1 to 690)		
COMMENT	Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P. The libraries that made SUCEST Genet. Mol. Biol. 24 (1-4), 1-7 (2001) Contact: Arruda P Centro de Biologia Molecular e Engenhariaia Genetica Universidade Estadual de Campinas Caixa Postal 6010, 13083-970, Campinas SP, Brazil Tel: 55 19 3788 1137 Fax: 55 19 3788 1089 Email: parruda@unicamp.br Clone distribution: clone name distribution information can be found through the Brazilian Clone Collection Center (BCCC) at <a href="http://www.bcccenter.fcav.unesp.br">http://www.bcccenter.fcav.unesp.br</a> Plate: 043 row: C column: 05 Seq primer: T7 Promoter Primer. Location/Qualifiers 1 .690 /organism="Saccharum officinarum" /mol_type="mRNA" /db_xref="taxon:4547" /clone="SCQGR1043C05" /lab_host="DH10B" /clone_lib="RT1" /note="Organ: Root tips (0.3cm-long) from adult plants; Vector: psport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [Root tips (0.3cm-long) from adult plants]. cDNA was prepared from polyA+ mRNA using SuperScript plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <a href="http://sucest.iad.ic.unicamp.br/public">http://sucest.iad.ic.unicamp.br/public</a> "		
FEATURES	source		
ORIGIN			
Query Match	2.9%;	Score 42.6;	DB 13; Length 690;
Best Local Similarity	51.7%;	Pred. No. 5.6;	
Matches	93;	Conservative	Mismatches 87; Indels 0; Gaps 0;
QY	1224	GAAATAATGTGGCCGGCGGCTCATGCTTTTGTTGAACCTCGACTGTGTTCGATCAGTATT	1283
Db	609	GAATNACGTGCGCGGAGAAGCTGGTACTTGNTGGACACCTCCCTGAAACCGATGAGCTGT	550
QY	1284	CCGAGCTTTTCGGGCAATACCAGGTGAACGCTGGCGATCTACGGCGGCTTCCGTTCT	1343
Db	549	CGGANGCGTGTTCGAGCGGTGAACGGGTGACCGCTGGTGGTGCAGCACCGCTTCCCGTCAT	490
QY	1344	CTGTCCGAGGACCTTAATCTCTTCGTAAAGTGCTTCCGATGGCTCGCTGATCAGGA	1403
Db	489	CGCGGTGTGACCGGACACCCAGAGCTGCACAGAGACGTGCGCGTCTCGTCTCGTGTCCCGA	430

```

RESULT 6
CAL131027/c
LOCUS
DEFINITION SCBFRT1067D04.g RT1 Saccharum officinarum cDNA clone SCBFRT1067D04
5', mRNA sequence.
ACCESSION
VERSION CAL131027
KEYWORDS
SOURCE
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 718)
REFERENCE
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
CONTACT: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parnuda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 067 row: D column: 04
Seq primer: T7 Promoter Primer.
FEATURES
Location/Qualifiers
1..718
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCBFRT1067D04"
/lab_host="DH10B"
/clone_lib="RTL"
/note="Organ: Root tips (0.3cm-long) from adult plants;
Vector: pSport1; Site 1: SalI; Site 2: NotI; An
unidirectional cDNA library generated from [root tips
(0.3cm-long) from adult plants]. cDNA was prepared from
polyA+ mRNA using SuperScript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://sucst.lad.ic.unicamp.br/public"
ORIGIN
Query Match 2.8%; Score 41.6; DB 13; Length 718;
Best Local Similarity 52.3%; Pred. No. 11;
Matches 92; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 1228 AATGTGGCCGGCCTATGCTTTGGTTGAACCTGACCTGTGTGGATCAGTATTCCGA 1287
Db AATGTGGCCGGCCTATGCTTTGGTTGAACCTGACCTGTGTGGATCAGTATTCCGA 1287
QY 605 AACGTGGCCGAGACCTGTACTTGTGACACCTCCCTGACCGATGAGGCTGTGGA 546
Db AACGTGGCCGAGACCTGTACTTGTGACACCTCCCTGACCGATGAGGCTGTGGA 546
QY 1288 GCCTTTTCCGGGCATACCCAGGTGAACGCTGGGATCTACGGCGCTTCGGTTCCTCTGT 1347
Db GCCTTTTCCGGGCATACCCAGGTGAACGCTGGGATCTACGGCGCTTCGGTTCCTCTGT 1347
QY 545 GCGGTGGTCGAGCGTGAACGGCTGACCGCTGGTGGTGCAGCACCAGCTTCCCGTCATCGG 486
Db GCGGTGGTCGAGCGTGAACGGCTGACCGCTGGTGGTGCAGCACCAGCTTCCCGTCATCGG 486
QY 1348 CCGGAGGACCTAATTTCTCGCTAAGGTCGTTCGATGGCTGCCTGATCAGGA 1403
Db CCGGAGGACCTAATTTCTCGCTAAGGTCGTTCGATGGCTGCCTGATCAGGA 1403
QY 485 GTTGACCGGCACCCAGAGCTGCAACGAGGACGTCGCGCTGCTGCGGTGCCGGA 430
Db GTTGACCGGCACCCAGAGCTGCAACGAGGACGTCGCGCTGCTGCGGTGCCGGA 430
RESULT 7
CNS017SY/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
1 (bases 1 to 1201)
REFERENCE
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)

```

```

ACCESSION AL108460
VERSION AL108460.1
KEYWORDS GI:5628764
SOURCE GSS
ORGANISM Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
Genoscope.
Direct Submission
AUTHORS Genoscope - Centre National de Sequencage :
TITLE BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
JOURNAL - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -- This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN37L08"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : Sp6"
ORIGIN
Query Match 2.8%; Score 41.4; DB 29; Length 1101;
Best Local Similarity 16.9%; Pred. No. 15;
Matches 41; Conservative 111; Mismatches 91; Indels 0; Gaps 0;
QY 70 GCGGTGTAGGCGAGTTCTTCACTCCATGAAGCGCCGACGCTGATGGCTTCAATGCTT 129
Db SSSASBSSSSSSBSSTSTTBTSTBBSSTSSSSSSSSBSBTSBSBTSBTTTBTBK 897
QY 130 CGGTGTGATGATCTCGGCGGACGCTGGGTGCTGACCCAGGAGCTGGTGGGCTCT 189
Db STSSTSSSTSBTTTBSBSSSSSSSSBSSTTTSBTSBSSYSSSBFTSKCSBTBSG 837
QY 190 CTGACCGTGCCTCGTGCATCGGCTGATCACTGAACGCCCGACGCTGCGGTCCACGTA 249
Db STSTBMSKBSSTSTSTBTSBSSBGGSCGSCGTSBSSCSBCKSKTSSTSGTTCGBC 777
QY 250 GTTGCCGTGGAACCGACCCCTTTTGTGCTTACCTGCGCCGACCCCTCGAGGATGT 309
Db SSGSCCCYCCCTCTCTCTCCSCGCTGSKBCKCYSTSSASTGSSCSSTSS 717
QY 310 CGG 312
Db SSG 714

```

```

RESULT 8
BX381961
LOCUS
DEFINITION Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSOD1072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
REFERENCE
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization

```



[illegible]

Db 147 CAGCCCGGTGTCCCGTATCTCGCGGATCGCTCGGTGCGCGCAGCGTCAATCCAGGCG 206  
 Qy 230 CCGAGCTTGGTCCAGTACGTTCCTGCGTGAACCGACCCCTTTGTCTGCTTACCTGC 289  
 Db 207 TGCAGGTGCGCGTGGCGTGAAGTCCGCGAGCGATCCGAATCCGCTCGGTACCGCC 266  
 Qy 290 GCGCCACCTCGAGGAATGTCCGAACGCTTACGGCATCTCTACGACCTAGTTCGAGGCG 349  
 Db 267 CGCCCGCTCAAAACCGAGGTGCGACGTGNNNGAGATATCCGCCACCACTCTGTCGGCG 326  
 Qy 350 ACTA 353  
 Db 327 CCTA 330

RESULT 10  
 BZ563177/c  
 LOCUS  
 DEFINITION pacs2-164\_4126.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
 BZ563177\_4126, genomic survey sequence.

ACCESSION BZ563177.1 GI:27186541  
 VERSION BZ563177.1  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 770)  
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

source  
 1..770

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pacs2-164\_4126"

/clone\_lib="pacs2-164"

/note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN

Query Match 2.8%; Score 41.2; DB 28; Length 770;  
 Best Local Similarity 50.5%; Pred. No. 14;  
 Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 129 TCGGTCGATGATCTCCGCGAAGCTGCGGTGCTCGACCCAGGAGCTGCTCGGTC 188  
 Db 341 TCAGCGCAGGTCCTCCGCGAGGCGCTGAATCCATGCCACGCGAGCTTCGCCGATCA 282

Qy 189 TCTGACCGCTGCGCTCGTGCATCGGTGATGAAACGCGCCGAGCTTGCCTCCACGT 248

Db 281 TCTCGTGAACCGCGCCCGGATGATGACGCGGACCTGCTGCTCGGGGATCG 222

Qy 249 AGTTCGCGTGAACCGACCCCTTGTGCTGCTTACCTGCGGCGCACCTCGAGGAATG 308

Db 221 AGAGGATCTTGATGAAGCCTTCGCTCTCGTGGTGTATCTTCGCCCGGCTGTTGGCGCTGA 162

Qy 309 TCGGAACGCTTACGGCAT 326

Db 161 ACGGGAAGCGTCCGACT 144

RESULT 11

BZ574354/c

LOCUS

DEFINITION

msb2\_363.y2 msh Pseudomonas aeruginosa genomic clone msh2\_363,

genomic survey sequence.

ACCESSION BZ574354

VERSION BZ574354.1

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1347)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

source  
 1..1347

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="MSH"

/db\_xref="taxon:287"

/clone="msh2\_363"

/clone\_lib="msh"

/note="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 2.8%; Score 41.2; DB 28; Length 1347;  
 Best Local Similarity 50.5%; Pred. No. 19;  
 Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 129 TCGGTCGATGATCTCCGCGAAGCTGCGGTGCTCGACCCAGGAGCTGCTCGGTC 188

Db 653 TCAGCACGAGGTCTCTCCGCGAGGCGCTGAATCCATGCCACGCGAGCTTCGCCGATCA 594

Qy 189 TCTGACCGCTGCGCTCGTGCATCGGTGATGAAACGCGCCGAGCTTGCCTCCACGT 248

Db 593 TCTCGCTGAACCGCGCCCGGATGATGTCACCGGACCTGCTGCTCGGGGATCGG 534

Qy 249 AGTTCGCGTGAACCGACCCCTTGTGCTGCTTACCTGCGCGCACCTCGAGGAATG 308

Db 533 AGAGGATCTTGATGAAGCCTTCGCTCTCGTGGTGTATCTTCGCCCGGCTGTTGGCGCTGA 474

Qy 309 TCGGAACGCTTACGGCAT 326

Db 473 ACGGGAAGCGTCCGACCT 456

RESULT 12

BM712537

LOCUS

DEFINITION

UI-E-EJO-ahh-k-14-0-UI.r1 UI-E-EJO Homo sapiens cDNA clone

UI-E-EJO-ahh-k-14-0-UI 5', mRNA sequence.

ACCESSION BM712537

VERSION BM712537.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 579)

BZ574354 1347 bp DNA linear GSS 17-DEC-2002  
 msb2\_363.y2 msh Pseudomonas aeruginosa genomic clone msh2\_363,  
 genomic survey sequence.

ACCESSION BZ574354

VERSION BZ574354.1

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1347)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

source  
 1..1347

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="MSH"

/db\_xref="taxon:287"

/clone="msh2\_363"

/clone\_lib="msh"

/note="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 2.8%; Score 41.2; DB 28; Length 1347;  
 Best Local Similarity 50.5%; Pred. No. 19;  
 Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 129 TCGGTCGATGATCTCCGCGAAGCTGCGGTGCTCGACCCAGGAGCTGCTCGGTC 188

Db 653 TCAGCACGAGGTCTCTCCGCGAGGCGCTGAATCCATGCCACGCGAGCTTCGCCGATCA 594

Qy 189 TCTGACCGCTGCGCTCGTGCATCGGTGATGAAACGCGCCGAGCTTGCCTCCACGT 248

Db 593 TCTCGCTGAACCGCGCCCGGATGATGTCACCGGACCTGCTGCTCGGGGATCGG 534

Qy 249 AGTTCGCGTGAACCGACCCCTTGTGCTGCTTACCTGCGCGCACCTCGAGGAATG 308

Db 533 AGAGGATCTTGATGAAGCCTTCGCTCTCGTGGTGTATCTTCGCCCGGCTGTTGGCGCTGA 474

Qy 309 TCGGAACGCTTACGGCAT 326

Db 473 ACGGGAAGCGTCCGACCT 456

BM712537 579 bp mRNA linear EST 28-FEB-2002  
 UI-E-EJO-ahh-k-14-0-UI.r1 UI-E-EJO Homo sapiens cDNA clone  
 UI-E-EJO-ahh-k-14-0-UI 5', mRNA sequence.

ACCESSION BM712537

VERSION BM712537.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 579)

**AUTHORS** Bonaldo,M.F., Lennon,G. and Soares,M.B.  
**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery  
**JOURNAL** Genome Res. 6 (9), 791-806 (1996)  
**MEDLINE** 97044477  
**PUBMED** 8889548  
**COMMENT** Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

**FEATURES** source  
 1..579  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-EJ0-ahh-k-14-0-UI"  
 /tissue\_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"  
 /dev\_stage="fetal and adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-EJ0"  
 /note="Organ: eye; Vector: pUT3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pUT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAAGT; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

**ORIGIN**  
 Query Match 2.8%; Score 41; DB 12; Length 579;  
 Best Local Similarity 51.4%; Pred. No. 14;  
 Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 169 CCAGGAGCGTGTGGGTCTCTGACCGCTGCGTGTGATCGGTGCTGATCGACGC 228  
 |||||  
 Db 146 CCAGGCCCGGTGTCGGTATCTCGGGATCCGCTGGTGGCGGAGCGTATCCAGGC 205  
 QY 229 CCGGAGCTTGGGTCCACGTAGTTCGGTGGAAACCGACCCCTTTGTCGTGCTTACCTG 288  
 |||||  
 Db 206 GTGCAGGTGCGGTGGTGGTGAAGTGGCGAGCGATCCGAATCCGGTGGGTACCGC 265  
 QY 289 CGCGCCACCTGAGGAGTATCGAAGCTTACGGCATCTCTACGACCTAGTCGAGGGC 348  
 |||||  
 Db 266 CGCGCGCTCAAAACCCAGGTTGGCGACGTGGTTCGAGATATCGCCACCACTCGTCGGC 325  
 QY 349 GACTA 353  
 |||||  
 Db 326 GCCTA 330

RESULT 13

BJ469381/c  
**LOCUS**  
**DEFINITION**

BJ469381 596 bp mRNA linear EST 23-MAY-2002  
 heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA  
 clone baal20k01 5', mRNA sequence.

**ACCESSION** BJ469381  
**VERSION** BJ469381.1 GI:21147884  
**KEYWORDS** EST.  
**SOURCE** Hordeum vulgare subsp. vulgare  
**ORGANISM** Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poideae; Triticeae; Hordeum.

1 (bases 1 to 596)

Sato,K., Saisho,D. and Takeda,K.  
 Barley EST sequencing project in NIG and Okayama Univ  
 Unpublished (2002)

**JOURNAL** Contact: Tadasu Shin-i

**COMMENT** Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

**FEATURES** Location/Qualifiers

source

1..596

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Haruna Niho"

/sub\_species="vulgare"

/db\_xref="taxon:112509"

/clone="baal20k01"

/tissue\_type="top three leaves"

/dev\_stage="adult, heading stage"

/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna

Niho adult, heading stage top three leaves"

**ORIGIN**

Query Match 2.8%; Score 41; DB 12; Length 596;

Best Local Similarity 52.7%; Pred. No. 14;

Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 124 ATCTTCGGGTGCGATCTCCCGGAACGTGCGGTGCTCCAGCCAGAGTGTGTC 183

|||||

Db 197 ATCTTCGGGCGGCTCGTGTGCGAGAGCGGTACTCGTGCATGACCGAGTTGGTGGTC 138

|||||

QY 184 GGGTCTCTGACCGCTGCCCTCGTCGATCGGCTGCATCTGAACGCCGCCGACGTTGGGTC 243

|||||

Db 137 GCCCTTGGGAGCGCGCGCGGTGGAAGACGACGCTCTTCTTCATCCGACGAGCGGCC 78

|||||

QY 244 CAGGTAGTTCGCGTGGAAACCGACCCCTTTGTCGTGCTTACCTGGCG 292

|||||

Db 77 GCGGTGGGACACCTCGCGGTCTTTCGCGGGTCTTCCAGTAGCGCGCG 29

|||||

**RESULT 14**

BJ472412/c

**LOCUS**

**DEFINITION**

BJ472412 661 bp mRNA linear EST 23-MAY-2002  
 heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA  
 clone baal34e02 5', mRNA sequence.

**ACCESSION** BJ472412

**VERSION** BJ472412.1 GI:21150915

**KEYWORDS** EST.

**SOURCE** Hordeum vulgare subsp. vulgare

**ORGANISM** Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Hordeum.

1 (bases 1 to 661)

Sato,K., Saisho,D. and Takeda,K.

Barley EST sequencing project in NIG and Okayama Univ

Unpublished (2002)

COMMENT  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
1. .661  
/organism="Hordium vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Haruna Nijo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baal34e02"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna  
Nijo adult, heading stage top three leaves"

FEATURES  
source  
1. .661  
/organism="Hordium vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Haruna Nijo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baal34e02"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna  
Nijo adult, heading stage top three leaves"

ORIGIN  
Query Match 2.8%; Score 41; DB 12; Length 661;  
Best Local Similarity 52.7%; Pred. No. 15;  
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 124 ATGCTTCGGTGCATGATCTCGCGAAGCGTGGCGGTCTCGACCCAGGAGCTGGTCTC 183  
Db 196 ATCTTGGGGCGCGTCTGTCAGAGAGCGGTACTGTCGTCATGACCCAGTGGTGGTCTC 137  
QY 184 GGGTCTCTGACCGTCCCTCGTCGATCGGTGTCATGACCGCCCGACGTTCCGGTCTC 243  
Db 136 GCCCTTGGAGCGCGCGCGGTGGAAGACGAGCGTCTTCTTCATCCCGACGAGCGCGC 77  
QY 244 CAGTAGTTCGGTGGAAACGACCCCTTTCGTCGCTTACCTACCTGGCG 292  
Db 76 GCGGTGGGACACCTCGCGTCTTTCGCGGGTCTTCCAGTAGCGCGCG 28

Search completed: October 2, 2004, 05:12:39  
Job time : 4258.29 secs

RESULT 15  
BX356664  
LOCUS  
DEFINITION  
BX356664 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1015YB03 3-PRIME, mRNA sequence.  
ACCESSION  
BX356664  
VERSION  
BX356664.1 GI:30378083  
KEYWORDS  
ESI.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D1015CA02NP1.

FEATURES  
source  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1015YB03"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 19:25:51 ; Search time 5949,09 Seconds  
(without alignments)  
10644.353 Million cell updates/sec

Title: US-10-668-047-1  
Perfect score: 1461  
Sequence: 1 gtgcatccgatcgccagcac.....ttccggaatcgctcgtga 1461

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_man.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	72.2	4.9	3889	1	PROIRM	K02081 P.stuartii
C 2	59.4	4.1	343250	1	AF0003594	AF0003594 Nostoc sp
C 3	51.2	3.5	1600	1	AF146609	AF146609 Aeromonas
C 4	49.6	3.4	2000	6	AX655393	AX655393 Sequence
C 5	45	3.1	2892	1	BACHSMR	L01541 Bacillus su
C 6	45	3.1	10929	1	U67554	U67554 Methanococ
C 7	45	3.1	110000	6	AR271569_04	Continuation (5 of
C 8	44.6	3.1	5039	1	YEN14030	AJ414030 Yersinia
C 9	44.2	3.0	1164	8	AK108161	AK108161 Oryza sat
C 10	44.2	3.0	146585	8	CNS08CAM	AL831809 Oryza sat
C 11	43	2.9	300900	1	AP005939	AP005939 Bradyrhiz
C 12	42.8	2.9	17786	1	RLHUPGENE	X52974 R.leguminos
C 13	42.8	2.9	21376	12	AF246703	AF246703 Synthetic
C 14	42.8	2.9	175376	2	AC108223	AC108223 Oryza sat
C 15	42.4	2.9	125020	9	AF429315	AF429315 Homo sapi
C 16	42.2	2.9	125020	9	AF429315	AF429315 Homo sapi
C 17	42	2.9	12337	1	AE011861	AE011861 Xanthomon
C 18	41.6	2.8	12046	1	AE005017	AE005017 Halobacte
C 19	41.2	2.8	11735	1	AE004896	AE004896 Pseudomon
C 20	41.2	2.8	12766	1	RLHUPREG	Z36981 R.leguminos
C 21	41.2	2.8	113193	1	AF357202	AF357202 Streptomy
C 22	41.2	2.8	113193	6	AX703543	AX703543 Sequence
C 23	41.2	2.8	132544	1	AF521085	AF521085 Streptomy
C 24	41.2	2.8	299925	1	AP005045	AP005045 Streptomy
C 25	41	2.8	299950	1	AP005963	AP005963 Bradyrhiz
C 26	40.8	2.8	302178	1	AE016918	AE016918 Chromobac
C 27	40.6	2.8	13993	1	AE001890	AE001890 Deinococc
C 28	40	2.7	302578	1	AE017208	AE017208 Geobacter
C 29	39.8	2.7	10393	1	AE011651	AE011651 Xanthomon
C 30	39.8	2.7	13325	1	AE005016	AE005016 Halobacte
C 31	39.8	2.7	219986	2	AC113026	AC113026 Mus muscu
C 32	39.8	2.7	299925	1	AP005045	AP005045 Streptomy
C 33	39.6	2.7	2000	6	AX655393	AX655393 Sequence
C 34	39.6	2.7	5364	1	SGKSACPG	X77865 S.griseus g
C 35	39.6	2.7	11847	1	AE001918	AE001918 Deinococc
C 36	39.6	2.7	300140	1	AP005956	AP005956 Bradyrhiz
C 37	39.4	2.7	12208	1	SGPREPG	X80774 Streptomyce
C 38	39.4	2.7	302101	1	AE016784	AE016784 Pseudomon
C 39	39.2	2.7	131676	9	AC139749	AC139749 Homo sapi
C 40	39.2	2.7	160766	2	AC137686	AC137686 Homo sapi
C 41	39	2.7	975	6	AX652811	AX652811 Sequence
C 42	39	2.7	1182	8	AK065644	AK065644 Oryza sat
C 43	39	2.7	25150	1	AB070945	AB070945 Streptomy
C 44	39	2.7	107870	8	AC105262	AC105262 Oryza sat
C 45	39	2.7	133271	8	AC104277	AC104277 Oryza sat

# ALIGNMENTS

RESULT 1  
PROIRM/c 3889 bp DNA linear BCT 16-FEB-1994  
LOCUS P.stuartii PstI restriction and modification genes, complete.  
DEFINITION K02081  
ACCESSION K02081.1 GI:150922  
VERSION endonuclease; methylase; modification enzyme; restriction  
KEYWORDS endonuclease.  
SOURCE Providencia stuartii  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Providencia.  
REFERENCE 1 (bases 1 to 3889)  
AUTHORS Walder,R.Y., Walder,J.A. and Donelson,J.E.  
TITLE The organization and complete nucleotide sequence of the PstI

Pred. No. is the number of results predicted by chance to have a



Kishida, Y., Kohara, M., Matsumoto, M., Matsumoto, A., Muraki, A., Nakazaki, N., Shimpō, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M. and Tabata, S.  
Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120  
Cyanobacteria 8 (5), 205-213 (2001)

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

Direct Submission  
Submitted (02-MAY-2001) Takazawa Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/cyanobase/  
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

## FEATURES

## source

1. .343250  
/organism="Nostoc sp. PCC 7120"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:103690"  
/note="synonym: Anabaena sp. PCC7120"  
complement (188. .6253)  
/gene="all3691"  
complement (188. .6253)  
/gene="all3691"  
/note="ORF\_ID: all3691"  
/codon\_start=1  
/transl\_table=11  
/product="serine/threonine kinase with two-component sensor domain"

## gene

## CDS

/protein\_id="BAB75390.1"  
/db\_xref="GI:17132825"  
/translation="MESVMNTYVDLAIAPGYFILEIIEYHSGKTVVRANREAOQOP  
WIKLLREYFTSLLQFRNQAIAKNIKPIGVNLSLEPINSYAMVEDFGIS  
LRDYAQOSSLTEILTIQADILHHYQRIIHKDIKPAIILINPETKQVLIID  
STASLLPRETONIISPNIQLEGTAVLSPQETGRMKRGDFSYSLGVSPELLTGE  
LPFTTDDMELVHCHIAKOPDKFTIPSCQRYANRSGRSTKLPEETIPQVGE  
IYMKLMKNAEDRYQNALGKLYDKLAELOETGKVFYITIGSRDVCDFIIPDKLY  
GREAEITLAAFRNVSIGATELMSVAGSGIGKTAIVNEVHKPIVRQGYEIQGKFD  
QQRNIPTSAFQRLMGLTSDTIQQRSKILBAVGKNGQVIEIPELELI  
IGKPAAPKLSGTSIQFNLLPKFKFTVFTSAEHLPLVFLDDQLWADSGSLKLIQLL  
MNDTGVLIIAGYRNEVNPAPLMTVNEIGKNGTTIKIDLA PLNOLKNTLVLET  
LKCSSETAQSLALVFOKQNGPEFTTFLKALHODNLIQFNYDLGCMQCDIAOVNHQ  
ALTDVILFMAQSCLKRLRATQEVQLACIGNQDFLELAIWESSPKHTAACLWAK  
LQEGLIIPQSEYKFFVQGEIHDQAQTEIVTYFLDRVQQAAYCLIPAEARAHK  
HHIGQLIKTSIPAAEEHIFIVNQLNYGITLI TQOERDELAQLNNAQKARAT  
AYQASREYADTSMILLGETAWQRYQLTITHELAAEIALCGDVEQNMQLLETVNH  
AKCPDRVOYQVKIQAITSRNELLEAIATGSKFLQELGVSLPDHPEDVQOARBEI  
NRLIGDRLGFEINLPKMTDPKLAIMOISSLIIPACVMTGSLVLLIIPLOVKLSIQ  
FGNSLPSAHGYVSVAFLSTTQWMAQAQQLQIAIYQALASEPAKNIRAAATFVLGGY  
FYHCTSHQDITLPIIOBQCIAGETGNLEFFGYNVQNGNAYSGESSEFASQMT  
YHQQLDNLVTSAHCFYVWESALITLGKSADEISLRQDAYAEKIVTEALANDLR  
LFQFYLRVLNFIADPRADKQAKQYLGCMGSAEPIFYFYDSLTALAEHQ  
PAGROSVORVRENALIDKADHAPMNVLHKWLVAAETHRLLOKTEAKETNDLEKR  
KGAKENKILDEALANEAALAEVLDWKEKVAAYVQRAYCYAREAKAKTNDLEKR  
YPLLQVLIQERQAFENFQETIAVPGTSSSTLSSTIGTSISDTDLFASILKAAQVIS  
TSLEDELIITNTEIILENSGAKSALLPQEDIQWIKAMTILSNFOPNSSESQTILE  
SQPLETCEIDIPKNIIVYKNTQVTDNLTQDIFGLIGYMLQHQPSQVFTPMWQ  
GHVGLTILENLTREGVFTSRLEVRILLSAQAVSLKARLYQESQTKAQQLKQSLK  
QOKILFNVNMQRSQSLDMAI FCVVTONIRRLIDVDVGVYQFHLVDVNYGEYFVAED  
VSPAPSAVAVQDPCGENYANLYKQGI CAITDVOSSEILDCRHOILAQFHVRA  
LVVPIQBEELMGLCICHCDDPRQWPELEMFAQVQAQMGIALKQTDLLIQTQQA  
TOLEHTLQHQLOQLQVQNERKAGNLVAGVHEMNNPLGFIASVLSQHTQPTFADI  
MEHLKLYOELPHGPKLHLHAAQIDLDILEDPLKIPAMVACDLKNI STLSRT  
SPADYDKYKIPNIGIDISTILIKHLRKANDKRAIEVITYEGLPHIECPGQOLN  
VPMNLANADAEDANNGLSYAEITANNRIIIRTTQVNNHVRISITDNGPMSSEL  
KQHIPEHLFTTKGVEKGTGLGLAIARQIWEKHGGAIEVNSQLQSGTEFVILPITAE  
SANKY"  
6804. .7658  
/gene="all3692"  
6804. .7658

## gene

## CDS

/gene="all3692"  
/note="ORF\_ID: all3692"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75391.1"  
/db\_xref="GI:17132826"  
/translation="MNFPQIGQVALALLTQGLKVLGAIALIWIQAOKLIEFALKLVR  
GLRTONIEPTLISLLNIVSVLRIVLIVAILGFFGIETTSFAALLAAGVAIGAAGW  
GILANFAAGAFLLVIRPEKVGDFITAAGITGTVSEIGLFTTITTPDNVLTIVANKI  
FADNIQNSANPFRFVLDLAQHDVDNQAIALKARISQIPNVNIENPAPDVEILTF  
NLGAPVLAVRYPYCNNDHYQVYFDTNKAIREFGEAGYPIPEQRYGFSVPENGTSSV  
IPSSALSS"  
8079. .8939  
/gene="all3693"  
8079. .8939  
/gene="all3693"  
/note="ORF\_ID: all3693"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75392.1"  
/db\_xref="GI:17132827"  
/translation="MNVAEISQVALALLTQGLRVLGAIALWIQAOLINFLGLKLLR  
AFRSKHVEPTLITNYIVNIVDVLKIVLIVAILGFFGIETTSFAALLAAGVAIGAAGW  
GILANFAAGAFLLVIRPEKVGDTIKAAVGTGKVEIGLFTTITSDNVKTIIGNNKI  
FADNIQNSANPFRFVLDLAQHDVDNQAIALKARISQIPNVNIENPAPDVEILTF  
NLLGTVLAVRYPYCHHEHYQVYFDTNKAIREFGEAGYPIPEQRYAFTGQSPDGTSA  
VQRTSLGRG"  
9664. .9831  
/gene="all3694"  
9664. .9831  
/gene="all3694"  
/note="ORF\_ID: all3694"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75393.1"  
/db\_xref="GI:17132828"  
/translation="MAVDLDKRLVLILQSFPHLRSWERNGLWEVSRRIKXETHVSM  
VAPEIELVIPG"  
complement (9841. .11007)  
/gene="all3695"  
complement (9841. .11007)  
/gene="all3695"  
/note="ORF\_ID: all3695"  
/codon\_start=1  
/transl\_table=11  
/product="N-acyl-D-glucosamine 2-epimerase"  
/protein\_id="BAB75394.1"  
/db\_xref="GI:17132829"  
/translation="MGNKLQAALYKNAALLNDVLFPWENHSLDTEGGYFTCLDRQK  
VYDIDKFIWLQNRQVTFMLCNQLEKKENMLKIASNGAKFLAQHGRDAGNWIYALT  
RGGEVLQFYNI FSDCFAMAFSOVALASGEAWKAVAMQYNNRKNPKGYTK  
TYPTGRKHALAPMILANTLMEWLLPQETLENVLSATVQEVMSDFLDKRGRLMYE  
NVAPGSHIDCFEGRILNPGHICIBAMWFTMDIARKNDSKTIQAVDVVNLTLNPAWD  
KEYGGLYVPMQAAGHPQQLWDOKLWHLSESLVALAMGYELTGRCVCMWYQKMD  
YSWQIPADPEYGEWGYLNRGRGEVLNLKGGKWKCGFHVPRALYLCWQOQFEALS"  
complement (11134. .12273)  
/gene="all3696"  
complement (11134. .12273)  
/gene="all3696"  
/note="ORF\_ID: all3696"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75395.1"  
/db\_xref="GI:17132830"  
/translation="MLPQALDNGRLLVAENGLTQOVLIVYKIQDQPVQVCTFGHTG  
GIYGVGEVQDLKFGYGTGVGAQAGNLIYNNNGNSGTDLRKFSFGKQLWLLG  
LSLVENADAPKTDGREFVTKYQFLIDSSKPKQQLIYKAYTLNGLFKYPQDSRLHIS  
PDTTFVRINDERFLFLDPSNALEIYRPNPKDGNIAIPAGMEVGNIRGKSAISG  
TWPPQEPGTGKIWRDNGDGAFDKEEYDRGEDSSSVTGWWYDSKGDVWTTILGDRRG

## gene

## CDS



gene  
RHYPLQGLDNGNPIYVINSIQEKTPLIFTLRRRIKYFPESDTMVLGFTAKNPPFA  
EDQAPGSEIACFDNWSKDNRIILRWIRVVPKDIRKEMITASH"  
complement(12278..13114)  
/gene="all13697"  
CDS  
complement(12278..13114)  
/gene="all13697"  
/note="ORF ID:all13697  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75396.1"  
/db\_xref="GI:17132831"  
/translation="MKHEILDEFKRLQSRRAKOSNKIFKVLVRLKSWIFLIIAIS  
FLAIIITADLTLSVAENSDDKLDGDIADSKSIQNIYEGIVYVAPNGTIVTNSHE  
DKAGAEASIVKDGNGVIGVSLDQMSRGGKAVTANSKYLIVAMSQGFVKIDKDP  
EGTWYCVRYNLSGRPAPFANGRGWDSMLIVNTKSEVTGLATKDKNLIVSDAANK  
IRVYNTETMELNRFSLRPEDITIDKQNLWIIQKODANNPAKILRYSSGKQLPOK  
NN"

gene  
14028..15197

/gene="hepB"

CDS  
14028..15197

/gene="hepB"

/note="ORF ID:alr3698

hepB gene product"

/codon\_start=1

/transl\_table=11

/product="heterocyst envelope polysaccharide synthesis

protein"

/protein\_id="BAB75397.1"

/db\_xref="GI:17132832"

/translation="MEYKREKLIVSASILTGLGWPTNPGLERYIYELTYQLSAN  
QDVELCGVLQPNQOHLPIKLTNLASPSDKIWOFWSIRNNFKTRIGKPDANLHF  
ALYSFPLDILPOGIPITFNPHGPWASESKOELVKNKISIFELKRLIETQYVNHCRF  
IVLSKAFNHLQOYQIPWPKIHIIPGVNIDKPNLSRQQAQRLQNWSPRILFT  
SRILRHGVNKLQALAIKPKLPDIWLAIGRHLOTTLEKQAQELGNVNFUG  
FLPDEQLPIAYAAANTVMPSPQFEGFLAITESLACGLVCLTPIGMEPIILTFSP  
QLITASPEATAIAEKIAQILLETQPKPSRECRQYAVTNFDWQKIAQVQRVILIA"  
15259..16407  
/gene="alr3699"  
15259..16407  
/gene="alr3699"  
/note="ORF ID:alr3699  
similar to glycosyltransferase"

Query Match 4.18; Score 59.4; DB 1; Length 343250;

Best Local Similarity 48.08; Pred. No. 0.002;  
Matches 270; Conservative 0; Mismatches 281; Indels 12; Gaps 3;

Qy 865 TTACTCTCGCTGATCGGCATCGGTGAAGTACGGGAAAGGTTGTGATTTTCGCAAT 924  
Db 17439 TCTACCTTAGATGAACITGGTTAGCAGTGTCAACGGGGCCAGTTGTAGATTTTCGCCTC 17498  
Qy 925 CGTCAGTATTGACCGATAACCTGGATGCTTCAGCGGTTGTGCCCATGTTTATCAGTCA 984  
Db 17499 AAATCAGCTTTGAGAAATATGTAATG---AGGAAACTGTTCATCTACTTACCCAGAA 17555  
Qy 985 AACATTCGATCTGTAATAATTGATTTGGCTTCAGTGGTGGAGGAAAGCCCTCAGGATTT 1044  
Db 17556 GCATATAAACAGGAAAGTTTATTTCCGCCAAAAAGCGCGTATAGTCATCGCCATT 17615  
Qy 1045 GTTCCGGTCGAAGATGTAGCATTACGTCAACTTCTCCCGCAAGGTCGTATGTTGTTGTG 1104  
Db 17616 GTACAAAATCAAGAAACAGACAAATGGCTAAT---CCATCAGGTGGTATGTTTTCACG 17672  
Qy 1105 AAAGCGCAACGGGAAAGGACCGCTGCTGCTGTGTCATCGCTGGGCTCG---GAC 1158  
Db 17673 AAGCGATTTTCTGTAAGAAGAAAAACCGCGGCTGTTGCTGCTGTATGTTCTTCCTGTA 17732  
Qy 1159 GGGCGCAGCAGGGTGGCTCGCAATAAAGCAACTATTGTCATGAATCTCAACGACCG 1218  
Db 17733 GATGCACCGAGTGTGGGCATAGAAATCACTCACTACTACCAATCTCAAGGTGAGGGA 17792  
Qy 1219 CTTGAGAAAAATGGCGCGGCTCATGCTTTGGTTGAACTCGACTGTTGGATCAG 1278

Db 17793 ATGAATCCCGACCTGGCAACGAGGTTTAGCAGCATTTCTCAGTTCAACTTTATTGGATAGT 17852  
Qy 1279 TATTTCCGAGCCTTTTCCGGGCATACCCAGGTGAACCTGGCGATCTACGCCGCTTCCG 1338  
Db 17853 TACTTCCGCTATTTCAGTGGACATACACAAGTAAATCAACGGATTTGCGGTAGAAATTAAG 17912  
Qy 1339 TTCTCTGTGCGGAGGACCTAAATTTCTCGTAAAGTGTTCCTCGATGGCCTGCCTGAT 1398  
Db 17913 TATCTTGCAGAGATGATTTAATCAAGCTAGTAGTCAAAATGGTGAATCTTCTGCTGAT 17972  
Qy 1399 CAGGAGAGCTTGGATGCCGTGT 1421  
Db 17973 CAGGCTCACTTGATGACGGTTGT 17995

RESULT 3

AF146609

LOCUS

DEFINITION

Aeromonas hydrophila putative modification methyltransferase gene,

partial cds.

ACCESSION

AF146609

VERSION

AF146609.2

GI:7327640

KEYWORDS

SOURCE

ORGANISM

Aeromonas hydrophila

Aeromonas hydrophila

Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;

Aeromonadaceae; Aeromonas.

REFERENCE

1 (bases 1 to 1600)

Zhang, Y.L., Ong, C.T. and Leung, K.Y.

Authors

10784058

REFERENCE

2 (bases 1 to 1600)

Zhang, Y.L., Ong, C.T. and Leung, K.Y.

Direct Submission

Submitted (28-APR-1999)

Department of Biological Sciences, National

University of Singapore, Singapore 119260, Singapore

Sequence update by submitter

On Mar 24, 2000 this sequence version replaced gi:7288241.

Location/Qualifiers

1..1600

/organism="Aeromonas hydrophila"

/mol\_type="genomic DNA"

/strain="PPD134/91"

/db\_xref="taxon:644"

185..>1600

/codon\_start=1

/transl\_table=11

/product="putative modification methyltransferase"

/protein\_id="AAF45040.2"

/db\_xref="GI:7327641"

/translation="MSSLPENIKHVNLLDPGCGSLTAFAVDEAIRGLTSLSH

AHDIEERTFPIDKTLQCMQACTNAGIDATKFLYNDFLINKTSDSKNVSKYTHC

IMNPYKKTASSSEHRRKRLHLAGTETVNLVAAFVLSVONLESQGLVAILPRSCNG

PYQPPRELILRETSIKHIIFESRNNAFSDDDVLOENIIHLIKNKQSDVITTSPP

IADPHDDDTGCVASDQKQFIIHANNRQELIDRLNSFA

TLDELNIQVSTGVPVDFRLKEDLRQNIPEGAVELIYFVHLNNGSWPKDKKPNAMV

SEQSKWLKWHGHGVIVRRFSKKEKRIYATNYSCLPGLIGENKLVFNHANKS

GIDRLAFGLTIYLNSTLDKYRLLFGGHTQVNNATDLRNHIHYPISNLSRMGSDVNL

DLTKQDIDGLDKELDMTGEN"

ORIGIN

Query Match 3.58; Score 51.2; DB 1; Length 1600;

Best Local Similarity 55.74; Pred. No. 0.13;

Matches 98; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1172 TTGGCTCGACAATAAAACGAACACTATTTGCATGATCTCAACGACCGCTTGAGAAAAATG 1231  
 Db 1317 TTGGCTTTGAAAATAAGCTTAATGTTATTTCAATGCAATAAGAGTGAATAGACCGAGATT 1376

QY 1232 TGGCCCGCGGCTCATGCTTTGGTTGAACCTGCATGCTGTTGGATCATGATTTCCGAGCCT 1291  
 Db 1377 TGGCCTTTGGTTTAAACGATTTATTTGAATCTTCACTCTTAGATAAAATATTACAGACTTT 1436

QY 1292 TTTCCGGGCATACCCAGGTGAACGCTGGCGATCTAGCCGCTTCCGTTCTCTCTGT 1347  
 Db 1437 TTGGTGGTCATACACAAGTAACGCTACAGACCTGAGAAATATTCAATCCCAAAGT 1492

RESULT 4  
 AX655393 2000 bp DNA linear PAT 22-MAR-2003  
 LOCUS Sequence 5263 from Patent WO03000898.  
 DEFINITION AX655393  
 ACCESSION AX655393  
 VERSION AX655393.1 GI:29158207  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa  
 Oryza sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1  
 AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,  
 Katagiri, F., Qian, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.  
 TITLE Plant genes involved in defense against pathogens  
 JOURNAL Patent: WO 03000898-A 5263 03-JAN-2003;  
 Syngenta Participations AG (CH)

FEATURES  
 source Location/Qualifiers  
 1..2000  
 /organism="Oryza sativa"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:4530"

ORIGIN

Query Match 3.4%; Score 49.6; DB 6; Length 2000;  
 Best Local Similarity 8.4%; Pred. No. 0.33;  
 Matches 75; Conservative 395; Mismatches 414; Indels 4; Gaps 1;

QY 144 CCGGGAACGGTGGGTGCTGACACCGAGAGTGTGTGGCTCTCTGACGCTGCCCT 203  
 Db 29 CARMGSRMSRKGMSKYRKCSSCKGKTKTRKSKWYSASSASGRTSKWSSGSYSGK 88

QY 204 CGTCGATCGCTGCATCTGAACGCCCGAGCTGGTGGTCCAGTGTGCCGTGGAAC 263  
 Db 89 GMKKRYKRKRMRGRGRGMRRSRMRMRGRRRCARSGRMAGSGRMGSGKSRMSYMWMC 148

QY 264 CGACCCCTTTGCTGCTTACCTGCGCGCACCTGGAGGATGTCGGNACGCTTACGG 323  
 Db 149 YARGCSCRKKSXGSGWGKTCRRGARGSGWSGAKYKSGSKRMKMWSSCGRSGCGR 208

QY 324 CACTCTACGACCTAGTCTGAGCGGCACTATTTGCTTTAAACGAGGCGCAAGCTTGATCG 383  
 Db 209 SAYSRVYGRSKYGYTKWYTSASRCMRAYMTTSYSWACSSYTWCRSKRSMWMMWMMRK 268

QY 384 CCGCTTCGATCTTGTAATGCTAATCTCCCTACGAAAGCTTGCTTCAGATTGCTGGC 443  
 Db 269 MRWSRSGYGWYSWYKWMWCTAYKYSYRWYCWYRGGWGRGATRYWGRGYMSRMAMMYKK 328

QY 444 CGCGCTTGCAACACAGCGGCTGCTGATGTACCGAAGCTTTACGTGGCCTCTGGGT 503  
 Db 329 MYWYRGYKMGKRWAGRMMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSRMSR 388

QY 504 CGCAGCAGTCATTTCGCTCAACAGCAGCGGGGGGGGTTTTTCATTGTTCTCGATCTTG 563  
 Db 389 KWCMRKRYKMRGYSRMSRKRWRWKKRSRGRWMMGRGOMTCRMKYSYGMWRKWSK 448

QY 564 GCGCAACGGCGCTTACTATCGTCAATTTCCGCTATTCGCTGATGACCGCGGTGAAGTCTCGA 623

[illegible]

```
/translation="WTOILETVDKSRILTVPNLLKNSKSELQGFQFTTSSISIFMACLFSE
DKLNNKVLADAGAGISLTSFAFLARLISENIGKADLHLLEIDEMLEPYLGETLALFKD
YIEINSQILIDDFIENWAXSLIDDESLANDKOPFTWAILNPPVKLIKSKHKKLKL
KAGLETNLYSAFVALVDLSDGGEIVFIIPRSFCNGPFRHFRHLLNKTSLKNNH
LFESRDKAFKNDVLENIKSKGKTVQBDVKISISTDDSFVSIRYRPFKEIVGP
NDIEKFTHINTNEETLIEKHNPVICYLEELNIEVSTGPPVDFRKNLREMPGEIVP
PLVPTNEFVGTSLSEYPMQMKPNAIIRNEKVKWLYPNKHVYVVKRFSKKEERIRIVA
GLVTPESVNDPVGVFENGVLNHYNKSGISKEVAYGLVAYLNSPVDKYRIFRNGHTQ
VNATDLRTMFPSRDILISLGKVIENIENVQGVIEDSKLEELLNDRGNA"
1697. .1701
/gene="hsmBI"
1714. .2664
/gene="hsmBI"
1714. .2664
/gene="hsmBI"
/codon_start=1
/transl_table=11
/product="restriction endonuclease"
/protein_id="AAAL18170.1"
/db_xref="GI:143055"
```

RBS

Gene

CDS

```
/translation="WTRGMSHNVEATIKELGLPKGOONERSALCLLSLMMITQDK
TWSAESPLGITPMFECRINYKEVAPNSREFRPTMHQFVDAGIALVNPDKPTR
PVNSKAVYQIEATLIEKYNTEESSELLARLYLSNQTILVERYAKERQONKIPVQI
AEGKEIYTPGEHELKAITIEEFAPRYVPGRLTYAGDTGKMGYFDEELRLQGVV
IDSHGKMPDVVYYPPEKKWLLILIESVTSHPGDKRHEELAKLFGNGSTAGIVYTAP
NRSLLMARYLNNISWETEVWVADAPSHLHFNGVRFLGPE"
```

ORIGIN

Query Match 3.1%; Score 45; DB 1; Length 2892;

Best Local Similarity 49.5%; Pred. No. 4.7;

Matches 154; Conservative 0; Mismatches 145; Indels 12; Gaps 1;

```
Qy 1075 CTTCTCCGCGAGGTCGTATGCTGTTGTAACCGGCAAGGCGAAGAGGACCGTGGT 1134
Dy 1332 CTTTACCTAATGGGCAITATGTTGTTTAAAGGTTTCTTCAAAGAGAAAACGT 1391
Qy 1135 CGTGTCACTCGTCGGTCT-----GGGACGGGGCGGACGAGGTTGCGTCGAC 1182
Dy 1392 CGAATAGTAGGAGGATTAACACGAGAGGTGTAACGATCCAGTAGTTGGTTTGAG 1451
Qy 1183 AATAAAACGAATATTGCAATGATCTCAACGACCGCTTGAGAAAATGTGGCCCGCGC 1242
Dy 1452 AACGACACTCAATGTACTTCACTATAATAAGAGTGGAAATTTCAAAGAAAGTAGCGTATG 1511
Qy 1243 CTCATGCTTGTGTTCACTCGACTGTGTTGATAGTATTTCCGAGCGCTTTCCGGGAT 1302
Dy 1512 CTTTATGCGTATCTTAACTCGACCCAGTAGATAGTATTCAGAAATTTTAATGTCAT 1571
Qy 1303 ACCGAGTGAACGCTGGCGATCTACCGCGCTTCCGTTCTCTGTCGCGAGGACCTAAT 1362
Dy 1572 ACTCAGGTAATAGCTACTGACCTTAGAACAAATGAAATTTCTTAGTAGAGATATCTAATA 1631
Qy 1363 CTTTCGCTAA 1373
Dy 1632 AGTTAGGTAA 1642
```

RESULT 6

U67554/c

LOCUS

DEFINITION

U67554 L77117

ACCESSION

U67554.1 GI:2826365

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

Merrick, J.M., Glodek, A., Scott, J.D., Geoghagen, N.S., Weidman, J.F., Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Hurst, M.A., Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C.  
Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
Science 273 (5278), 1058-1073 (1996)  
96337999  
8688087  
2 (bases 1 to 10929)  
Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weidman, J.F., Merrick, J.M., Glodek, A., Scott, J.D., Geoghagen, N.S., Weidman, J.F., Fuhrmann, J.L., Nguyen, D.T., Utterback, T., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Hurst, M.A., Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C.  
Direct Submission  
Submitted (27-AUG-1996) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
On Jan 30, 1998 this sequence version replaced gi:1591751.

Location/Qualifiers

1. .10929

/organism="Methanocaldococcus jannaschii"

/mol\_type="genomic DNA"

/db\_xref="taxon:2190"

180..710

/gene="MJ1109"

180..710

/note="MJ1109"

/note="similar to SP:P09556 percent identity: 38.37;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="orotate phosphoribosyl transferase (pyrE)"

/protein\_id="AAB99112.1"

/db\_xref="GI:1592320"

/translation="WDKSKLNLKLVGCFRFGFIFILASGKSNYYIDIKATTNPE

ILKLVGEIIASQIKEDVDKAGVELGVSPIATAVSIIAQPLLIVRKPKDVGTKNKI

EGLKRGKVVIVEDVTGGSVLKAVKEIRENGIVDKVFFVVDLEGAKENLQKEN

VELPLVTVKELQSTQ"

complement (808..2268)

/gene="MJ1110"

complement (808..2268)

/note="similar to GB:AE000666 percent identity: 30.00;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAB99113.1"

/db\_xref="GI:1499961"

/translation="WESTNLKIANEFDSLNEINDKLKEVVGKKNKIDTKLSD

IILTEKLEILQDLKMEFLFDSYKVNKKGKGGYDSEGLQETASTYLRRIAS

BKGLERVRHALVAHIALAHLTDIGNINLPNPLDGSYKIMFEPFVLYVTYK

RFLDLEPKRGILTSYISLIVIDKREPKRIVKDEKVEKYEKKEFGNAIITSIK

RNFKNKIIDQYVVRVAIGYLVNTYKDEIKAEIKNEIKDLNEEBEKYINRYELCL

LFREBADSGLLDVRCMEERKLELEKEILKGLYDGEPIELPKKAIKRNELS

KKISKDIIKRFSEVDVFKYLTTPDRARSNLPSISMITPQRFSLWSVDGINVD

VLDFKPLEELPKYQIFLKNITGGVALYIHDWDAVERNFNFKKDIEDLLKIALIEP

IKELDKNDVDVSKLEKFGKVKETKFLDLLSL"

complement (2231..3580)

/gene="MJ1111"

complement (2231..3580)

/note="similar to GP:1575544 percent identity: 38.14;

identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="tRNA nucleotidyltransferase (cca)"





QY 387 GTTCGATCTTTGTAATGCTATCTCCTCCTACGAAAGCTTGCTTCAGATTGCTGGCGG 446  
 Db 1501 GTACACACATTTGTAATTAATCCCTTACAAAAATAACATCTGCAAGTATTACCG 1560  
 QY 447 GCTTGCACACACGCGGTGCGTGCATGATCCGAACTTTACGTGGCTTCTGGGTGCG 506  
 Db 1561 TAAATATTAAGTCTATCGGTATTAAGCTGTTAATCTATATCTGCTGTTTGTGCTCT 1620  
 QY 507 AGCAGTCAATTTTCGCTCAAGAGCAGCGCGGGGTTTTTCATGCTTCTCGATCTTGGC 566  
 Db 1621 TGCTATCATGCAATTAATAAACHAGGTGAATGCTAGCAATAATCCAGATCTTTTGG 1680  
 QY 567 GAACGGCCTTATCTATCGTCAATTCGCCATTTGGCTGATGACCGCGGTGAAGTCTTCGATAT 626  
 Db 1681 TAAATGGCCATATTAATCTACCTTCAGGAATTTTATATTTTTCAGCACTGTGCCATCAACA 1740  
 QY 627 ACTTCATGTTGTCGAAAGTAGAACCAAGATATTTGGCGACGAAAGTAAGCAAGAA 686  
 Db 1741 TGTTCATATATTCGATAGTGAAGTCATGCAATCTCTGAAGATGATGTACTCCAAGAAA 1800  
 QY 687 TGTTCATGTTGCTTTTCAGTGTGA 709  
 Db 1801 TATCATTAATTAATGTTAAGA 1823

## RESULT 9

AKI08161/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

## REFERENCE

AUTHORS

1 The Rice Full-length cDNA Consortium, National Institute of Agrobiological Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ohtsuki, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Nariyama, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayashida, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.  
 Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
 Science 301 (5631), 376-379 (2003)  
 2752273  
 12869764

## TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 1164)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, J., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,

Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ohtsuki, K., Ohtsuki, K., Saitoh, H., Sakai, C., Sakai, K., Sakagawa, A., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami, T., Tagami, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.  
 Direct Submission  
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp)  
 Tel: 81-29-838-7007, Fax: 81-29-838-7007  
 This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/  
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.  
 FAIS Genome Sequencing & Analysis Group: Ohtsuki, J., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariyama, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

Location/Qualifiers  
 1. 1164  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="002-139-G12"

## FEATURES

source

1. 1164

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="mRNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="002-139-G12"

ORIGIN

Query Match

Best Local Similarity

Matches

88; Conservative

0; Mismatches

73; Indels

0; Gaps

0;

136

GATGATCTCGCGAACGGTGGGTGCTCGACCCAGGAGCTGGTTCGACCTCTGACC

195

830

GAGGAGCGCGCGTGTGAGGAGGGTGTGAAGACCCAGGCGCGATCGCGTGCACCGCC

771

196

GCTGCCCTCGTCATCGGCTGCTGATCTGAACGCCGAGCTGCGGTTCACGTGTC

255

770

GCCGCGAGCGCGCCCGAAGTGCATCCACGACGCGCGAGCGGAGGTTGACCC

711

256

GTGGAACACCGACCCCTTTCTGCTGCTTACCTTCGCGGCCAC

296

710

GAGGACGCGCATCTGCACGCGCGAGCGCGGCGGCCAC

670

## RESULT 10

CNS08CAM

LOCUS

CNS08CAM

LOCUS

146585 bp

DNA

linear

PLN 21-NOV-2003







RLHUPGENE/c  
LOCUS RLHUPGENE 17786 bp DNA linear BCT 26-JAN-1996  
DEFINITION R.leguminosarum bv. viciae UPW791 hydrogenase cluster.  
ACCESSION X52974  
VERSION X52974.1 GI:1167855  
KEYWORDS hupC gene; hupD gene; hupE gene; HupE mature protein; hupF gene; hupG gene; hupH gene; hupI gene; hupJ gene; hupK gene; hupL gene; hupS gene; hydrogenase large subunit; hydrogenase small subunit; hupA gene; hupB gene; hupC gene; hupD gene; hupE gene; hupF gene; hupX gene; Pseudo-HoxA.  
SOURCE Rhizobium leguminosarum  
ORGANISM Rhizobium leguminosarum  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
REFERENCE 1 (bases 1 to 3052)  
AUTHORS Hidalgo,E., Leyva,A. and Ruiz-Argueso,T.  
TITLE Nucleotide sequence of the hydrogenase structural genes from Rhizobium leguminosarum  
JOURNAL Plant Mol. Biol. 15 (2), 367-370 (1990)  
MEDLINE 91355885  
PUBMED 2103457  
REFERENCE 2 (bases 3053 to 5595)  
AUTHORS Hidalgo,E., Palacios,J.M., Murillo,J. and Ruiz-Argueso,T.  
TITLE Nucleotide sequence and characterization of four additional genes of the hydrogenase structural operon from Rhizobium leguminosarum bv. viciae  
JOURNAL J. Bacteriol. 174 (12), 4130-4139 (1992)  
MEDLINE 92283771  
PUBMED 1597428  
REFERENCE 3 (bases 5596 to 8668)  
AUTHORS Rey,L., Hidalgo,E., Palacios,J. and Ruiz-Argueso,T.  
TITLE Nucleotide sequence and organization of an H2-uptake gene cluster from Rhizobium leguminosarum bv. viciae containing a rubredoxin-like gene and four additional open reading frames  
JOURNAL J. Mol. Biol. 228 (3), 998-1002 (1992)  
MEDLINE 93108466  
PUBMED 1469733  
REFERENCE 4 (bases 8669 to 14781)  
AUTHORS Rey,L., Murillo,J., Hernando,Y., Hidalgo,E., Cabrera,E., Imperial,J. and Ruiz-Argueso,T.  
TITLE Molecular analysis of a microaerobically induced operon required for hydrogenase synthesis in Rhizobium leguminosarum biovar viciae  
JOURNAL Mol. Microbiol. 8 (3), 471-481 (1993)  
MEDLINE 93316844  
PUBMED 8326860  
REFERENCE 5 (bases 1 to 17786)  
AUTHORS Imperial,J., Rey,L., Palacios,J.M. and Ruiz-Argueso,T.  
TITLE HupK, a hydrogenase-ancillary protein from Rhizobium leguminosarum, shares structural motifs with the large subunit of NifE hydrogenases and could be a scaffolding protein for hydrogenase metal cofactor assembly  
JOURNAL Mol. Microbiol. 9 (6), 1305-1306 (1993)  
MEDLINE 95020662  
PUBMED 7934943  
REFERENCE 6 (bases 1 to 17786)  
AUTHORS Brico,B., Palacios,J.M., Hidalgo,E., Imperial,J. and Ruiz-Argueso,T.  
TITLE Nickel availability to pea (Pisum sativum L.) plants limits hydrogenase activity of Rhizobium leguminosarum bv. viciae bacteroids by affecting the processing of the hydrogenase structural subunits  
JOURNAL J. Bacteriol. 176 (17), 5297-5303 (1994)  
MEDLINE 94350809  
PUBMED 8071205  
REFERENCE 7 (bases 1 to 17786)  
AUTHORS Rey,L., Imperial,J., Palacios,J.M. and Ruiz-Argueso,T.  
TITLE Purification of Rhizobium leguminosarum HupB, a nickel-binding protein required for hydrogenase synthesis  
JOURNAL J. Bacteriol. 176 (19), 6066-6073 (1994)  
MEDLINE 95014042  
PUBMED 7928968  
REFERENCE 8 (bases 14604 to 16400)  
AUTHORS Rey,L., Brito,B., Fernandez,D., Hernando,Y., Palacios,J., Imperial,J. and Ruiz-Argueso,T.  
TITLE The hup cluster from Rhizobium leguminosarum bv. viciae UPW791 contains an additional gene, hupX, required for nickel processing and hydrogenase maturation  
JOURNAL Unpublished  
REFERENCE 9 (bases 16401 to 17786)  
AUTHORS Fernandez,D., Rey,L., Palacios,J., Imperial,J. and Ruiz-Argueso,T.  
TITLE A pseudo-hoxA gene in the Rhizobium leguminosarum bv. viciae UPW791 hup cluster  
JOURNAL Unpublished  
REFERENCE 10  
AUTHORS Ruiz-Argueso,T.  
TITLE Direct Submission  
JOURNAL Submitted (10-MAY-1990) Ruiz-Argueso, T., E.T.S. Ingenieros Agronomos, Universidad Politecnica Madrid, Laboratorio de Microbiologia, E.T.S. Ingenieros Agronomos, 28040 Madrid, Spain  
REMARK Revised by [11]  
REFERENCE 11 (bases 1 to 17786)  
AUTHORS Rey,L.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-1996) Rey L., E.T.S. Ingenieros Agronomos, Universidad Politecnica Madrid, Laboratorio de Microbiologia, E.T.S. Ingenieros Agronomos, 28040 Madrid, Spain  
COMMENT On Jan 27, 1996 this sequence version replaced gi:48720.  
FEATURES  
source  
1..17786  
/organism="Rhizobium leguminosarum"  
/mol\_type="genomic DNA"  
/strain="UPW791"  
/db\_xref="taxon:384"  
misc\_feature  
1..14781  
/note="hups, hupL, hupC, hupD, hupE, hupF, hupG, hupH, hupI, hupJ, hupK, hupA, hupB, hupF, hupC, hupD, hupE hydrogen uptake genes"  
/label="hydrogen\_uptake"  
31..43  
/bound\_moiety="IHF consensus"  
misc\_binding  
145..150  
/label="PI\_promoter"  
156..1238  
/gene="hups"  
156..1238  
/gene="hups"  
promoter  
join(71..87,99)  
/note="71..87 NtrA-binding consensus sequence;99 Transcription initiation site"  
RBS  
145..150  
gene  
156..1238  
CDS  
156..1238  
/gene="hups"  
/function="hydrogenase small subunit precursor"  
/mat\_peptide  
291..1235  
/product="hydrogenase small subunit"  
/function="hydrogenase small subunit"  
/citation=[6]  
/evidence=experimental  
1262..3052  
/gene="hupL"  
1262..3052  
/gene="hupL"  
/function="hydrogenase large subunit precursor"

RESULT 13	AF246703	21376 bp	DNA	linear	STN 06-NOV-2000
LOCUS	Synthetic construct minitransposon TnHB100 sequence.				
DEFINITION	AF246703				
ACCESSION	AF246703.2 GI:11079645				
VERSION	synthetic construct				
KEYWORDS	artificial construct				
SOURCE	1 (bases 1 to 21376)				
ORGANISM	Bascones,E., Imperial,J., Ruiz-Argueso,T. and Palacios,J.M.				
REFERENCE	Generation of new hydrogen-recycling Rhizobiaceae strains by introduction of a novel hup minitransposon				
AUTHORS	Appl. Environ. Microbiol. 66 (10), 4292-4299 (2000)				
TITLE	20466836				
JOURNAL	MEDLINE				
MEDLINE	PUBMED				
REFERENCE	2 (bases 1 to 21376)				
AUTHORS	Bascones,E., Imperial,J., Ruiz-Argueso,T. and Palacios,J.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAR-2000) Biotechnology, Univ. Politecnica de Madrid, Ciudad Universitaria s/n, Madrid, Madrid 28040, Spain				
COMMENT	On Nov 3, 2000 this sequence version replaced gi:11034796.				
FEATURES	Location/Qualifiers				
	1..21376				
	/organism="synthetic construct"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32630"				
	/transposon="TnHB100"				
	/note="miniTn5-type transposon"				
repeat_region	1..19				
	/note="Tn5 left repeat"				
	/rpt_type=inverted				
misc_feature	complement(28..2063)				
	/note="Omega interposon"				
misc_feature	2224..3257				
	/note="corresponds to the Rhizobium leguminosarum bv. viciae putative 2-hydroxychromene-2-carboxylate isomerase and putative alcohol dehydrogenase genes in GenBank Accession Number U23040"				
misc_feature	3258..3383				
	/note="corresponds to the Rhizobium leguminosarum bv. viciae hupSL gene promoter in GenBank Accession Number U36924"				
misc_feature	3384..19767				
	/note="corresponds to Rhizobium leguminosarum sequence in GenBank Accession Number X52974"				
repeat_region	21358..21376				
	/note="Tn5 right repeat"				
	/rpt_type=inverted				
ORIGIN					
Query Match	2.9%; Score 42.8; DB 12; Length 21376;				
Best Local Similarity	58.7%; Pred. No. 20;				
Matches	74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;				
QY	178	GGTGTGGGTCCTGTGACGGTCCCTGTCGATCGGCTGCATCTACTGAACGCCCGACGTT	237		
Db	14029	GGCAATGGCTTCTGTGTCGGCGCTTGGCAGCCGCAACAGGCGCATCGCTT	13970		
QY	238	CGCGTCCACGTAGTTGCCGTGGAACCGACCCCTTTTGTGCTGCTTACCTGCGCGCCACC	297		
Db	13969	CGCGGCGTGCAGAAAGTGGAAACCGCGACGCCCTTGAGCGCAACGATCGCGCGCC	13910		
QY	298	CTGGAG 303			
Db	13909	TTCCGAG 13904			
RESULT 14	AC108223/c	175376 bp	DNA	linear	HTG 28-FEB-2003
LOCUS	AC108223				
DEFINITION	Oryza sativa (japonica cultivar-group chromosome 11 clone				

OSUNBa0019A16, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 5 ordered pieces.

AC108223  
 HTG; HTGS PHASE2.  
 Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartodeae; Oryzoae; Oryza.  
 1 (bases 1 to 175376)  
 Sharma,T.R., McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N.,  
 Kuit,K., Miller,B., Nascimento,L.U., O'Shaughnessy,A.L.,  
 Preston,R.R., Santos,L., Spiegel,L.A., Zutavern,T., Mohapatra,T.  
 and Singh,N.K.  
 Genomic sequence for Oryza sativa chromosome 11  
 Unpublished  
 2 (bases 1 to 175376)  
 Singh,N.K.  
 Direct Submission  
 Submitted (26-JAN-2002) Indian Initiative for Rice Genome  
 Sequencing, NRC Plant Biotechnology, Indian Agricultural Research  
 Institute, Pusa Road, New Delhi 110012, India  
 3 (bases 1 to 175376)  
 Sharma,T.R., McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N.,  
 Kuit,K., Miller,B., Nascimento,L.U., O'Shaughnessy,A.L.,  
 Preston,R.R., Santos,L., Spiegel,L.A., Zutavern,T., Mohapatra,T. and  
 Singh,N.K.  
 Direct Submission  
 Submitted (28-FEB-2003) IIRGS, NRC on Plant Biotechnology, Indian  
 Agricultural Research Institute, LBS Centre, New Delhi, Delhi  
 110012, India  
 On Jan 6, 2003 this sequence version replaced gi:20531974.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 1658: contig of 1658 bp in length  
 \* 1659 1758: gap of unknown length  
 \* 1759 58136: contig of 56378 bp in length  
 \* 58137 58236: gap of unknown length  
 \* 58237 72536: contig of 14300 bp in length  
 \* 72537 72637: gap of unknown length  
 \* 72637 104055: contig of 31419 bp in length  
 \* 104056 104155: gap of unknown length  
 \* 104156 175376: contig of 71221 bp in length.  
 repeat\_region Location/Qualifiers  
 1..175376  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:39947"  
 /chromosome="11"  
 /clone="OSUNBa0019A16"  
 /note="(japonica cultivar-group)"  
 ORIGIN  
 Query Match 2.9%; Score 42.8; DB 2; Length 175376;  
 Best Local Similarity 46.1%; Pred. No. 24;  
 Matches 143; Conservative 0; Mismatches 167; Indels 0; Gaps 0;  
 QY 34 CAGAGTGTCTCGGAACTGACCCCACTACTCAAGCGGTCTAGGCGAGTCTTCTACT 93  
 Db 147991 CAGGAGGTGGCCACCTCTCCACGCGTCTCTCTCGCGGTGGCGGTCTCGCTCGCG 147932  
 QY 94 CCCATGAGGCGGCCACGCTGATGGTTCATAGTCTCGGTCGATGATCTCCGCGGAACG 153  
 Db 147931 CGCGCGGCTGCTCAGCAGATCTCCCTGGCGCTCGCGCGCTCGCGCGCGAG 147872  
 QY 154 GTGCGGTGCTCGACCCAGGAGCTGGTGTCTGCGGTCTCTGACCGCTGCCCTCGATCGG 213

Db 147871 GGCGGGTGGACGGCTCTTTCGCGGATGGGCACCTCCCGACCCCGCGTGTGATGGAG 147812  
 QY 214 CTGCATACTGAACGCCCGGAGTTGCGGTCCACCTAGTTCGCTGGAAACGACCCCTTT 273  
 Db 147811 CTGCTCACTGTGCTCCCGAGGAGGTGTGTCAGGACGAGCGGGGACACCGGGTGGAC 147752  
 QY 274 GTCTGCTTACCTGCGGCCACCTCGAGGAATGTTCGAAACGCTTACGGCATCTCTTAC 333  
 Db 147751 GCCCGGGCGCTGCAGGTTACCCGCGAGTCTATGGCGCAGCGCCGCGTGTCTGAG 147692  
 QY 334 GACCTAGTGC 343  
 Db 147691 TTCTCTCTCG 147682

RESULT 15  
 AF429315 125020 bp DNA linear PRI 18-JAN-2002  
 LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.  
 DEFINITION  
 ACCESSION AF429315  
 VERSION AF429315.1 GI:17646244  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 125020)  
 Holmes,S.B., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,  
 Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,  
 Potter,N.T., Ross,C.A. and Margolis,R.L.  
 A repeat expansion in the gene encoding junctophilin-3 is  
 associated with Huntington disease-like 2  
 Nat. Genet. 29 (4), 377-378 (2001)  
 JOURNAL 2 (bases 1 to 125020)  
 MEDLINE 11694876  
 PUBMED  
 REFERENCE  
 AUTHORS Holmes,S.B., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical  
 Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
 FEATURES  
 source Location/Qualifiers  
 1..125020  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /map="16q24.3; between DL6S520 and WI-12410"  
 /note="Isolated from a patient with Huntington's  
 Disease-Like 2 (HDL2)"  
 complement(35581..35746)  
 /rpt\_type=tandem  
 /rpt\_unit="ctg"  
 complement(<36507..>36887)  
 /gene="JPH3"  
 /note="synonym: JP3"  
 complement(<36507..>36887)  
 /gene="JPH3"  
 /product="junctophilin 3"  
 complement(<36507..36887)  
 /gene="JPH3"  
 /note="Component of the junctional complex between plasma  
 membrane and endoplasmic reticulum"  
 /codon\_start=1  
 /product="junctophilin 3"  
 /protein\_id="AA140941.1"  
 /db\_xref="gi:17646245"  
 /translation="MSSGGRNFDDGSGYCGWEDGKAHGHCVTGPKGGRYTGSWS  
 HGFVILGYTFPSGNTYQGTWAQGRHGIGLESKGWYKGEWTHFGKRYGVRECA  
 NGAKYEGTWSNLQDGYGTETYSQ"  
 ORIGIN  
 Query Match 2.9%; Score 42.4; DB 9; Length 125020;



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 1, 2004, 19:21:23 ; Search time 625.714 Seconds  
(without alignments)  
9919.257 Million cell updates/sec

Title: US-10-668-047-1  
Perfect score: 1461  
Sequence: 1 gtgcacgcgacgcgcagcac.....ttccggaatctgctcgtga 1461

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002s:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	49.6	3.4	2000	7	ADA71938	Ada71938 Rice gene
2	45	3.1	110000	2	AAV21209_04	Continuation (5 of
3	44.8	3.1	536	9	ADB68842	Adb68842 Minority
4	42.2	2.9	1947	6	AA140127	Aal140127 Isoprenoi
5	41.2	2.8	1425	7	ACA26157	Aca26157 Prokaryot
6	41.2	2.8	113193	7	AAD54645	Aad54645 Streptomy
7	41	2.8	1311	7	ACA26275	Aca26275 Prokaryot
8	39.6	2.7	1503	7	ACA25789	Aca25789 Prokaryot
9	39.6	2.7	2000	7	ADA71938	Ada71938 Rice gene
10	39.4	2.7	987	7	ACA43826	Aca43826 Prokaryot
11	39.4	2.7	1119	7	ACA23836	Aca23836 Prokaryot
12	39	2.7	975	7	ADA69358	Ada69358 Rice gene
13	38.8	2.7	551	7	ABX74255	Abx74255 DNA seque
14	38.8	2.7	9810	2	AAZ32025	Aaz32025 Human MET
15	38.8	2.7	9810	5	AAZ32025	Aaz32025 Human MET
16	38	2.6	2000	9	AAAC90082	Aac90082 AF018073
17	37.6	2.6	633	8	ADA49246	Ada49246 Maize gen
18	37.6	2.6	110000	2	AZ01425_01	Continuation (2 of
19	37.6	2.6	110000	2	AZ01425_02	Continuation (3 of
20	37.4	2.6	789	6	ABN87153	Abn87153 Lolium pe
21	37.2	2.5	687	7	ACA38429	Aca38429 Prokaryot
22	37.2	2.5	690	7	ACA40639	Aca40639 Prokaryot
23	37.2	2.5	110000	4	AA199682_26	Continuation (27 o

24	37.2	2.5	110000	4	AA199683_25	Continuation (26 o
25	37.2	2.5	110000	4	AA199683_26	Continuation (27 o
26	36.8	2.5	925	2	AAV15076	Aav15076 Hybrid DN
27	36.8	2.5	38726	4	AAAS59513	Aas59513 Propionib
28	36.8	2.5	38726	7	ACF64442	Acf64442 Propionib
29	36.6	2.5	367	3	AAC56312	Aac56312 Pinus rad
30	36.6	2.5	3453	7	ACA26750	Aca26750 Prokaryot
31	36.4	2.5	1017	9	ADC23900	Adc23900 DNA seque
32	36.4	2.5	2522	6	ABX98456	Abx98456 Consensus
33	36.4	2.5	6291	4	ABL15665	Ab115665 Drosophil
34	36.4	2.5	14260	4	ABL15664	Ab115664 Drosophil
35	36.4	2.5	14615	4	AAAS59577	Aas59577 Propionib
36	36.4	2.5	14615	7	ACF64506	Acf64506 Propionib
37	36.4	2.5	30001	2	AA761016	Aat61016 Total DNA
38	36.4	2.5	30001	2	AA761016	Aat61016 Total DNA
39	36.2	2.5	788	6	ABN87154	Abn87154 Lolium pe
40	36.2	2.5	1233	7	ACA42148	Aca42148 Prokaryot
41	36.2	2.5	1395	4	AAAS1511	Aas1511 Pseudomon
42	36.2	2.5	1395	7	ACA19477	Aca19477 Prokaryot
43	36.2	2.5	1486	6	ABN87145	Abn87145 Lolium pe
44	36.2	2.5	2277	3	AAZ53328	Aaz53328 Neisseria
45	36.2	2.5	2295	3	AAZ53319	Aaz53319 Neisseria

ALIGNMENTS

RESULT 1  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.  
XX  
AC ADA71938;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 5263.  
XX  
KW Plant; bacterial infection; fungal infection; viral infection; rice;  
gene; ds.  
XX  
OS Oryza sativa.  
XX  
PN WO2003000898-A1.  
XX  
PD 03-JAN-2003.  
XX  
PF 22-JUN-2001; 2001WO-IB001105.  
XX  
PR 22-JUN-2001; 2001WO-IB001105.  
XX  
(SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
WPI; 2003-175290/17.  
XX  
PT Identifying at least one gene involved in plant resistance or response to  
pathogenic infection for conferring resistance or tolerance to a plant to  
bacterial, fungal or viral infection by determining or detecting plant  
gene expression.  
XX  
PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX  
The present invention relates to a method (M1) for identifying genes  
involved in plant resistance or response to pathogenic infection. M1  
comprises identifying a gene whose expression is significantly altered in  
the incompatible interaction of plant gene expression relative to  
the expression of the gene in an uninfected plant, in a mutant plant that  
does not express a gene associated with response to pathogenic infection,  
or in a corresponding incompatible or compatible interaction. (M1) is  
useful for conferring resistance to resistance or tolerance to a plant to  
bacterial, fungal or viral infection. The present sequence was used to







```
DR WPI; 2003-029926/02.
DR P-PSDB; ABU22287.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 14027; 1765pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1425 BP; 205 A; 499 C; 528 G; 193 T; 0 U; 0 Other;
Query Match 2.8%; Score 41.2; DB 7; Length 1425;
Best Local Similarity 52.3%; Pred. No. 0.12;
Matches 91; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 125 TGCTTCGGGTGCGATGATCTCCGCGAAGCGGTGCGGTGCTGACCCAGGAGCTGGTGTGC 184
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236 TGGCGCGCGTGCATCAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 177
QY 185 GGTTCTTGACCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 244
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
176 AGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 117
QY 245 ACCTAGTTCGCGGAAACGACCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 298
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 GCGACGATGGTGGCGCGGTATGCGACGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63
RESULT 6
AAD54645/c
ID AAD54645 standard; DNA; 113193 BP.
XX
AC AAD54645;
XX
DT 26-JUN-2003 (first entry)
XX
DE Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster..
XX Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX Streptomyces nodosus.
XX OS
```

```
XX Location/Qualifiers
FH complement(4. .1824)
FT
FT /*tag= a
FT /product= "ABC transporter encoded by S. nodosus amphG
FT gene"
FT complement(1805. .3628)
FT
FT /*tag= b
FT /product= "ABC transporter encoded by S. nodosus amphH
FT gene"
FT 3840. .4874
FT /*tag= c
FT /product= "GDP-mannose dehydratase encoded by S. nodosus
FT amphDIII gene"
FT 5042. .33574
FT /*tag= d
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT nodosus amphI gene"
FT 33584. .50518
FT /*tag= e
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 15, 16 and 17 encoded by S. nodosus
FT amphJ gene"
FT 50571. .56675
FT /*tag= f
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 18 and thioesterase encoded by S.
FT nodosus amphK gene"
FT 56829. .58019
FT /*tag= g
FT /product= "Cytochrome P450 encoded by S. nodosus amphL
FT gene"
FT 58139. .58648
FT /*tag= h
FT /product= "ORF1, hypothetical protein"
FT complement(58756. .59610)
FT /*tag= i
FT /product= "ORF2, hypothetical protein"
FT 59869. .61470
FT /*tag= j
FT /product= "ORF3, hypothetical protein"
FT /transl_except= (pos:59869. .59871, aa:Met)
FT complement(61798. .61995)
FT /*tag= k
FT /product= "Ferredoxin encoded by S. nodosus amphM gene"
FT complement(62051. .63250)
FT /*tag= l
FT /product= "Cytochrome P450 encoded by S. nodosus amphN
FT gene"
FT complement(63250. .64308)
FT /*tag= m
FT /product= "NDP-sugar aminotransferase encoded by S.
FT nodosus amphDII gene"
FT complement(64324. .65775)
FT /*tag= n
FT /product= "Glycosyl transferase encoded by S. nodosus
FT amphDI gene"
FT /transl_except= (pos:65773. .65775, aa:Met)
FT 66081. .70319
FT /*tag= o
FT /product= "Polyketide synthase multienzyme housing
FT loading module encoded by S. nodosus amphA gene"
FT 70366. .79938
FT /*tag= p
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 1 and 2 encoded by S. nodosus by amphB
FT gene"
FT 79956. .112709
FT /*tag= q
FT /product= "Polyketide synthase multienzyme housing
FT extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT nodosus by amphC gene"
FT
```

XX PN WO200297082-A2.  
XX PD 05-DEC-2002.  
XX FF 27-MAY-2002; 2002WO-IE000071.  
XX PR 31-MAY-2001; 2001IE-00000527.  
XX PA (UYDU-) UNIV COLLEGE DUBLIN.  
XX FI Caffrey JP;  
XX WPI; 2003-201271/19.  
DR P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,  
DR AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,  
DR AAE36129, AAE36130, AAE36131, AAE36132.  
XX PT Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful  
PT for preparing amphotericin derivative or analog antibiotic agent with  
PT altered properties, in biosynthesis of polyketide other than  
PT amphotericin.  
XX PS Claim 1; Page 52-114; 276pp; English.  
XX CC The invention relates to the gene cluster encoding the polypeptides  
CC responsible for the biosynthesis of the polyene antibiotic amphotericin  
CC (amph) of Streptomyces nodosus. Polynucleotides of the invention are  
CC useful for preparing amphotericin derivatives or analogue antibiotic  
CC agents with altered properties and in the biosynthesis of polyketides  
CC other than amphotericin. amphotericin, amphotericin or amphotericin are useful  
CC for producing amphotericin derivatives glycosylated with alternative  
CC sugars; amphotericin or amphotericin gene sequences are useful in engineered  
CC biosynthesis of perosaminyl-amphoteronolide B; amphotericin or amphotericin and  
CC amphotericin gene sequences are useful in the engineered biosynthesis of  
CC perosaminyl-16-desacetoxy-16-methyl amphoteronolide B; amphotericin, amphotericin  
CC and amphotericin gene sequences are useful for preparing polypeptides capable  
CC of addition of mycosamine to a polyketide other than amphoteronolide A or  
CC B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.  
CC The present sequence is S. nodosus amph biosynthetic gene cluster  
XX SQ Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;  
Query Match 2.8%; Score 41.2; DB 7; Length 113193;  
Best Local Similarity 48.7%; Pred. No. 1.2;  
Matches 112; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 127 CTTCCGGTTCGATGATCTCCGCGGAACCGGTGGTGTCTCGACCCGAGCTGGTCTCGGG 186  
DB 26692 CTCAGGAGGTGGTCTCTCCGAGCGGTCCGCGCGTGGAGCGCGAGCGTGGCAGG 26633  
QY 187 TCTCTGACCGTCCCTCGTTCGATCGGCTGCATCTGATCTGACCCCGAGTGGGTTCAC 246  
DB 26632 TGCTCTGCGCGAGCGGTCCGCGGTCTGCTGGCGCGCGGTGAGCTGTCTCGTCCGG 26573  
QY 247 GTAGTTCGCGTGAACCGACCCCTTTGTCTGCTTTACCTGCGCGCCACCTCGAGGAA 306  
DB 26572 ACCGATCTCGAGAGGTGGTGTACACCGTCTCTCGCGAGCGCCGCGAGCGGTCTCGGCGAA 26513  
QY 307 TGTGCGGAACCGTTCAGGCACTTCTCTACGACCTAGTTCGAGGCGCGACTATTT 356  
DB 26512 CCGTACCGTCTCGCGGAGCGTCCCGACCCAGTACCGCGGGGTGAGGAGTT 26463  
RESULT 7  
ACA26275  
ID ACA26275 standard; DNA; 1311 BP.  
XX AC ACA26275;  
XX AC ACA26275;  
DT 19-JUN-2003 (first entry)  
XX DE Prokaryotic essential gene #7932.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX FF drug design; gene.  
XX OS Burkholderia mallei.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-023926/02.  
DR P-PSDB; ABU22405.  
XX CC New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 14; SEQ ID NO 14145; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1311 BP; 162 A; 466 C; 489 G; 194 T; 0 U; 0 Other;  
Query Match 2.8%; Score 41; DB 7; Length 1311;  
Best Local Similarity 46.9%; Pred. No. 0.13;  
Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
QY 108 CAGCGCTGATGGCTTCAATGCTTCGGGTTCGATGATCTCCGCGGAACGCTGGGTCTCGA 167  
DB 696 CAGCGCGCTCGCGGATGGGTTCGCGCGGAGATGATGACGACGCGGTCTGCGGA 755



PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
PS Claim 27; SEQ ID NO 5263; 899pp; English.  
XX  
CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
Query Match 2.7%; Score 39.6; DB 7; Length 2000;  
Best Local Similarity 8.1%; Pred. No. 0.45;  
Matches 30; Conservative 179; Mismatches 163; Indels 0; Gaps 0;  
Qy 1078 CTCGCGAGGGTGTATGTTGTTGTAACGGCAAGAGGACCGTCGTCGT 1137  
Db 662 MKSCWYWRGARSWYKYSKSCAKCKKTRTYTSSYMSYGYSSYKSMSTKMSYM 603  
Qy 1138 GTCATCGTCGCGTCTGGAGCGGCCAGAGGGTTGGCTCGACAATAAAGCACTAT 1197  
Db 602 GKWTCTWYTSKGRSTRSKGRWSGMSRMYYRWKMRKRYMYMKWTWRRCMYRW 543  
Qy 1198 TTGCATGAATCTCAACGACCGCTTGAGAAATGTGGCCCGCCCTCATGCTTTGGTTG 1257  
Db 542 GYTYMTTSRSRMVYTGKRYKRYTSSKRYMYKRYCWYYYGYMYKSYMYRYCYKAC 483  
Qy 1258 AATCGAGTGTGGATGATGATTCGAGCCCTTTCCGGGGATACCAGGTGAAGCT 1317  
Db 482 KKCCYACWCAAYSGSMYMYRYKYSKWRMSYKYMWSMYKCRSMYKYGKCGKMT 423  
Qy 1318 GGCATCTACCGCGCTTCGCTCTCTGTGCGGAGGACCTAATCTCTCGCTAAGTC 1377  
Db 422 YCSGYMKWYTYMGSKYKRCYKTYRMWYKGMWYMYSAYSNMWYIYAKYKYYW 363  
Qy 1378 GTTCCCGATGCGCTGCTGATCAGAGACGTTGGATGCGGTGGCCAGACTCTTCTGT 1437  
Db 362 KRRGTMSWYKYSKKYCTWYCMKRCYXWRKMYKRYKRCYCWRYATCYWCCY 303  
Qy 1438 GAGATTCGGNA 1449  
Db 302 RKGWYSRRSNM 291  
RESULT 10  
ACA43826/c  
ID ACA43826 standard; DNA; 987 BP.  
XX  
AC ACA43826;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #25483.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Pseudomonas putida.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU39956.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 31696; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway; (8)  
CC required for proliferation, or that inhibits cellular proliferation; (9)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 987 BP; 200 A; 327 C; 301 G; 159 T; 0 U; 0 Other;  
Query Match 2.7%; Score 39.4; DB 7; Length 987;  
Best Local Similarity 54.5%; Pred. No. 0.36;  
Matches 79; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
Qy 134 TCATGATCTCCGCGGAACGGTGGGTGCTCGACCCAGAGCTGTGTGCGGTCTCTGA 193  
Db 584 TTGCTGAATCTGACGAAGCGCGGGCGCTGCTGCAGCAGCGTGTGCGCCAGGCTCG 525  
Qy 194 CCCTGCGCCCTCGTCGATCGGCTGCATCTGAAACGCCCCGACGTTGCGGTCCAGTAGTG 253  
Db 524 GTGCTGCTGTCGATCCAGTCAGCATGTTGCTGTTGTTGACCGCCCACTTG 465  
Qy 254 CCCTGGAACCCGACCCCTTTGTCGT 278  
Db 464 CCCTGCGGGTGCACCGCCATGCTT 440  
RESULT 11  
ACA23836/c  
ID ACA23836 standard; DNA; 1119 BP.

XX AC A23836;  
 XX 19-JUN-2003 (first entry)  
 XX DE Prokaryotic essential gene #5493.  
 XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX KW drug design; gene.  
 XX OS Borrelia cepacia.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX P-PSDB; ABU19966.  
 XX WPI; 2003-029926/02.  
 XX DR New antisense nucleic acids, useful for identifying proteins or screening  
 XX PT for homologous nucleic acids required for cellular proliferation to  
 XX PT isolate candidate molecules for rational drug discovery programs.  
 XX PS Claim 14; SEQ ID NO 11706; 1766pp; English.  
 XX CC The invention relates to an isolated nucleic acid comprising any one of  
 XX CC the 6213 antisense sequences given in the specification where expression  
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 XX CC encoding a polypeptide whose expression is inhibited by the antisense  
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 XX CC polypeptide or its fragment whose expression is inhibited by the  
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 XX CC proliferation or the activity of a gene in an operon required for  
 XX CC proliferation; (7) identifying a compound that influences the activity of  
 XX CC the gene product or that has an activity against a biological pathway  
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
 XX CC identifying a gene required for cellular proliferation or the biological  
 XX CC pathway in which a proliferation-required gene or its gene product lies  
 XX CC or a gene on which the test compound that inhibits proliferation of an  
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 XX CC compound's activity; (11) a culture comprising strains in which the gene  
 XX CC product is overexpressed or underexpressed; (12) determining the extent  
 XX CC to which each of the strains is present in a culture or collection of  
 XX CC strains; or (13) identifying the target of a compound that inhibits the  
 XX CC proliferation of an organism. The antisense nucleic acids are useful for  
 XX CC identifying proteins or screening for homologous nucleic acids required  
 XX CC for cellular proliferation to isolate candidate molecules for rational  
 XX CC drug discovery programs, or for screening homologous nucleic acids  
 XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 XX CC prokaryotic essential genes. Note: The sequence data for this patent did  
 XX CC not form part of the printed specification, but was obtained in  
 XX CC electronic format directly from WIPO at  
 XX CC ftp.wipo.int/pub/published\_pt\_sequences  
 XX SQ Sequence 1119 BP; 221 A; 382 C; 360 G; 156 T; 0 U; 0 Other;

Best Local Similarity 48.1%; Pred. No. 0.38;  
 Matches 112; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
 QY 90 CACTCCCATGAAGCGCGCCACCGCTGATGCTTCAATGCTTGGGTGATGATCTCGCGG 149  
 DB 862 CACCCGCGAAGCGACCGCTGTCGATACGCGCCCGTCCCTTCGCGACCGGAATCGTGT 803  
 QY 150 AACGGTGGGGTGCTCGACCCAGGAGCTGGTGTCCGGGTCTCTGACCGCTGCCCTCGTCA 209  
 DB 802 GCGCAGCAGGCTGTCGACGAGCCGCTGTCGCTTCGAACGCGAGTAGCCCGGTAGA 743  
 QY 210 TCGGCTGCATGTAACGCGCCCGACGTTGCGGTCCACGTAGTTCGCGTGGAAACCGACCC 269  
 DB 742 TCTTGGCTTGTGTAACGCGTACGATGTTCCGCTGGTAGATCGTCTCTTGGCGCGCG 683  
 QY 270 CTTTGTGTCGCTTACCTGCGCGCCACCTCGAGGAGATGTCGGAACGCTTACG 322  
 DB 682 AGTTGTGCTGTCTGCTGCGCGCGCGCGATGCTGAGCGGACCATCAG 630  
 RESULT 12  
 ADA69358/c  
 ID ADA69358 standard; DNA; 975 BP.  
 XX AC ADA69358;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Rice gene, SEQ ID 2681.  
 XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
 XX KW gene; ds.  
 XX OS Oryza sativa.  
 XX PN WO2003000898-A1.  
 XX PD 03-JAN-2003.  
 XX PF 22-JUN-2001; 2001WO-IB001105.  
 XX PR 22-JUN-2001; 2001WO-IB001105.  
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 XX PI Katagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX WPI; 2003-175290/17.  
 XX PT Identifying at least one gene involved in plant resistance or response to  
 XX PT pathogenic infection for conferring resistance or tolerance to a plant to  
 XX PT bacterial, fungal or viral infection by determining or detecting plant  
 XX PT gene expression.  
 XX PS Claim 6; SEQ ID NO 2681; 899pp; English.  
 XX CC The present invention relates to a method (M1) for identifying genes  
 XX CC involved in plant resistance or response to pathogenic infection. M1  
 XX CC comprises identifying a gene whose expression is significantly altered in  
 XX CC the incompatible interaction of plant gene expression relative to  
 XX CC expression of the gene in an uninfected plant, in a mutant plant that  
 XX CC does not express a gene associated with response to pathogenic infection,  
 XX CC or in a corresponding incompatible or compatible interaction. (M1) is  
 XX CC useful for conferring resistance to resistance or tolerance to a plant to  
 XX CC bacterial, fungal or viral infection. The present sequence was used to  
 XX CC illustrate the invention.  
 XX SQ Sequence 975 BP; 182 A; 364 C; 284 G; 145 T; 0 U; 0 Other;  
 Query Match 2.7%; Score 39; DB 7; Length 975;  
 Best Local Similarity 53.6%; Pred. No. 0.47;  
 Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;



Qy	118	GCTTCAATGCTTCGGGTCGATGATCTCCGCGGAACGGTCCGGTGTCTCGACCCAGGAGCT	177
Db	834	CGCGCACTGCTGGTTCGGCGCATCATCTCCGGGAGAGGCCCCAGCGCTGGAGCTCCTCCTT	775
Qy	178	GGTGTCCGGTCTCAGACCGCTCCCTCGTTCGATCGGCTGCATCTGAACGCCGCCGAGCTT	237
Db	774	GGTGTGGTGGATGAGGAGGCTGTACTCCGCGGGGCGGAGCTGCATGACCCCTGAACCTC	715
Qy	238	CGCGTCCACGCTAGTTGCCGTGGAACCCGACC	268
Db	714	GCAGTTGTCGAAGTTGTTGAGGCGCTCGNAC	684
RESULT 13			
ABX	74255	ID	ABX74255 standard; DNA; 551 BP.
XX		AC	ABX74255;
XX		DT	18-MAR-2003 (first entry)
XX		XX	DNA sequence #27 encoding M. tuberculosis antigenic polypeptide.
DE		XX	Mycobacterium tuberculosis antigenic polypeptide; immune response;
XX		XX	tuberculosis infection; gene therapy; gene; ds.
KW		XX	Mycobacterium tuberculosis.
OS		XX	US6456533-B1.
PN		XX	15-OCT-2002.
PD		XX	23-DEC-1999; 99US-00470191.
PF		XX	24-DEC-1998; 98US-0113952P.
PR		XX	(CORI-) CORIXA CORP.
PA		XX	Skeiky Y;
PI		XX	WPI; 2003-147072/14.
DR		XX	Novel isolated mycobacterial polynucleotide, useful for treating,
PT		XX	preventing or diagnosing Mycobacterium tuberculosis infection, for
PT		XX	producing Mycobacterium tuberculosis secretory polypeptides and DNA
PT		XX	vaccines.
XX		XX	Example; Col 45-46; 48pp; English.
XX		XX	The present invention relates to the isolation of polynucleotide
CC		XX	sequences encoding Mycobacterium tuberculosis antigenic polypeptides. The
CC		XX	polynucleotide sequences of the invention are useful for treating, M.
CC		XX	preventing, and diagnosing M. tuberculosis infection, for producing M.
CC		XX	tuberculosis secretory polypeptides, for producing DNA vaccines, for
CC		XX	diagnostic purposes, as molecular probes or primers to detect the
CC		XX	presence of bacteria in a biological sample, for inducing and/or
CC		XX	enhancing immune responses to M. tuberculosis, and in gene therapy.
CC		XX	ABX74229-ABX74319 represent DNA sequences encoding M. tuberculosis
CC		XX	antigenic polypeptides
SQ		XX	Sequence 551 BP; 87 A; 183 C; 172 G; 99 T; 0 U; 10 Other;
Query Match			
			2.7%; Score 38.8; DB 7; Length 551;
			Best Local Similarity 44.7%; Pred. No. 0.4;
			Matches 139; Conservative 0; Mismatches 172; Indels 0; Gaps 0
Qy	97	ATGAAGGCGCGCAGCTGATGGCTTCAATGGTTCGGGTTCGATCTCGCGGAACGGTG	156
Db	174	ATCCGCGGGGCGCANGCGTCGCAATGGTTCGATNGTGGCGCCACCGANGCTTG	233
Qy	157	CGGTGCTCGAACCGAGGAGCTGGTGTCCGGTCTCTGAACCGTGCCTCGTCGATCGGCTG	216

D	b		234	GGCCTGCGTACCGATCCGCCAGGGTCAGCGGCGGTACCACCCCAGCGCGTGCGCAGCGGCCG	239
Q	y		217	CATACTGAACGCCCCGACGTTGGCGGTCCA CGTAGTTGCGTGAAACCACCCCTTTGTC	276
D	b		294	CGGGTCCACCGCCGCAACTGACGTGCTTTGGCGGCGCGTGGCATCGTGAAGCATC	353
Q	y		277	GTGCTTACCTGCGCGCCACCCCTGGAGGAATTGTCGGAACGCTTA CCGCATCTCCTACGAC	336
D	b		354	GTCCGTGCGCGGGACACGCCGGAATTGGCAATCGMACCCCGGCATCCGAGCGATCGC	413
Q	y		337	CTAGTCGAGGGCGA CTATTTGCTTTAACCAAAGGGGCCAAGCTGGATGCCCGTTCGATCTT	396
D	b		414	CTCCGGGGTACA CAACAAGCGCAAGCGCGCGGGTTGCGGN TTGCGCNATTA AACTT	473
Q	y		397	GTAATTGCTAA 407	
D	b		474	GTTGTTGGCAA 484	
R	E	S	U	L	T
I	D	A	Z	3	2
X	X				
A	C				
X	X				
D	T				
X	X				
D	e				
K	w				
K	w				
K	w				
K	w				
K	w				
K	w				
O	s				
X	x				
P	n				
X	x				
p	d				
X	x				
P	f				
P	r				
P	r				
X	x				
F	a				
(HAST//	HASTINGS G A.				
(RUBE//)	RUBEN S M.				
Iruela-Arispe L,	Hastings GA, Ruben SM;				
WPI; 1999-590684/50.					
New isolated metalloproteinase thrombospondin polypeptides, useful for treating hyperproliferative disorders, cancers or autoimmune disorders.					
Disclosure; Page 353-359; 457pp; English..					
AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human metalloproteinase Thrombospondin (METH) proteins METH1 and METH2 respectively. METH1 and METH2 have been found to be potent inhibitors of angiogenesis both in vitro and in vivo. They can be used for treating cancer and other disorders related to angiogenesis including abnormal wound healing, inflammation, rheumatoid arthritis, psoriasis, endometrial bleeding disorders, diabetic retinopathy, some forms of macula degeneration, haemangiomas, and arterial-venous malformations. They may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilisation (chemotaxis) of immune cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some CC autoimmune disorders, acquired (e.g. by chemotherapy or toxins). Or					



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 23:40:18 ; Search time 116.496 Seconds  
(without alignments)  
6959.771 Million cell updates/sec

Title: US-10-668-047-1

Perfect score: 1461

Sequence: 1 gtgcattccgacgcagcac.....ttccggaatctgctctgta 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	3.1	1664976	4	US-08-916-421B-1
2	41.2	2.8	840	4	US-09-252-991A-10289
3	41.2	2.8	1260	4	US-09-252-991A-10493
C 4	41.2	2.8	1635	4	US-09-252-991A-10567
C 5	41.2	2.8	1743	4	US-09-252-991A-10674
C 6	38.8	2.7	551	4	US-09-470-191-27
7	38	2.6	507	4	US-09-252-991A-15931
8	38	2.6	1056	4	US-09-252-991A-15895
9	38	2.6	3357	4	US-09-252-991A-15868
C 10	37.8	2.6	1257	4	US-09-252-991A-4342
C 11	37.8	2.6	1353	4	US-09-252-991A-4513
C 12	37.8	2.6	1419	4	US-09-252-991A-15497
C 13	37.8	2.6	2016	4	US-09-252-991A-15291
C 14	37.8	2.6	2508	4	US-09-252-991A-15550
C 15	37.4	2.6	930	4	US-09-252-991A-4633
C 16	37.4	2.6	1179	4	US-09-252-991A-4241
C 17	37.4	2.6	1719	4	US-09-252-991A-4534
C 18	37.2	2.5	4403765	3	US-09-103-840A-2
C 19	37.2	2.5	4411529	3	US-09-103-840A-1
C 20	36.8	2.5	525	4	US-09-252-991A-14501
C 21	36.8	2.5	922	3	US-09-189-060B-73
C 22	36.8	2.5	1083	4	US-09-252-991A-3016
C 23	36.8	2.5	1092	4	US-09-252-991A-14759
C 24	36.8	2.5	1452	4	US-09-252-991A-2653
C 25	36.8	2.5	2145	4	US-09-252-991A-14638
C 26	36.8	2.5	2487	4	US-09-252-991A-14985
C 27	36.4	2.5	30001	1	US-08-125-468-1

C 28	36.4	2.5	30001	2	US-08-474-933-1	Sequence 1, Appli
C 29	36.2	2.5	1329	4	US-09-252-991A-15277	Sequence 15277, A
C 30	36.2	2.5	1551	4	US-09-252-991A-15327	Sequence 15327, A
C 31	36	2.5	1098	4	US-03-252-991A-4322	Sequence 4322, Ap
C 32	36	2.5	1137	4	US-09-252-991A-4155	Sequence 4155, Ap
C 33	36	2.5	3924	4	US-09-252-991A-4296	Sequence 4296, Ap
C 34	35.8	2.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 35	35.8	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 36	35.4	2.4	2634	3	US-08-941-936-1	Sequence 1, Appli
C 37	35.4	2.4	3120	4	US-09-221-017B-54	Sequence 54, Appli
C 38	35	2.4	645	4	US-09-252-991A-2971	Sequence 2971, Ap
C 39	35	2.4	678	4	US-09-252-991A-893	Sequence 893, App
C 40	35	2.4	1272	4	US-09-252-991A-16045	Sequence 16045, A
C 41	35	2.4	1575	4	US-09-252-991A-1139	Sequence 1139, Ap
C 42	35	2.4	1587	4	US-09-023-655-1192	Sequence 1192, Ap
C 43	35	2.4	1830	4	US-09-252-991A-3164	Sequence 3164, Ap
C 44	35	2.4	1980	4	US-09-252-991A-16555	Sequence 16555, A
C 45	35	2.4	2859	4	US-09-252-991A-944	Sequence 944, App

#### ALIGNMENTS

RESULT 1  
US-08-916-421B-1  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.  
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii  
; Patent No. 6503729  
; TITLE OF INVENTION: jannaschii  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/024,428  
; FILING DATE: 1996-08-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1664976  
; TYPE: DNA  
; ORGANISM: Methanococcus jannaschii  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (28222)..(28222)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (28257)..(28258)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84773)..(84773)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84808)..(84808)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84812)..(84812)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98120)..(98120)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98159)..(98159)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98239)..(98239)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98266)..(98266)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98343)..(98343)

```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191389)..(191389)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231380)..(231380)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
```

US-08-916-421B-1

Query Match 3.1%; Score 45; DB 4; Length 1664976;

Best Local Similarity 47.4%; Pred. No. 0.058;

Matches 135; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
QY 681 AGAGAAATGTCATCGTTGCTTTTCAGTGTGAGCGCGCAAGCTCTAGTGTGTCTCTCTAG 740
Db 416035 AGTTAATGTTTATCTGGTGTTCATTTACAGACCAAGGATTTATCGAAGGCAAGCTTA 416094
QY 741 GTCGGTCCGACATGGAGAGAGTCGATCGCAGTTCTGTGCGCTTTCTGCGCTTCTCA 800
Db 416095 TGAATTAGCCAAGCTAAAAAGTTGCTTCCAGATAAGATAAAGTTTTTTCAGATTTCA 416154
QY 801 TGATGAAGACGATGACAAAATCTGTCACITTCGCGGAAAGCGCATCGGTGCGCGCGG 860
Db 416155 TGTAAGCATGCATATCATTTTATAGACITTTGAAGCTCATTTGTGATACCGTTGAGAG 416214
QY 861 GAGGTTTACTCTCGCTGATCTCGCATCGGTGTAAGTACGGAAAGGTTGTTGATTTTCG 920
```

Db 416215 AGGTTAGCTGATGCTGTAATATATCAGCGGTAAAGAGAACGGGAAAGAGGTTGATATTGA 416274  
QY 921 CAATCGTCAGTATTGACCGATAACCTGGATGCTTCAGCGGTTGT 965  
Db 416275 AAAGCTAAATAGCTAAAGGAATTGGTTGATGTTCCAGTTATTGT 416319

## RESULT 2

US-09-252-991A-10289  
; Sequence 10289, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10289  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10289

Query Match 2.8%; Score 41.2; DB 4; Length 840;  
Best Local Similarity 50.5%; Pred. No. 0.019;  
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 129 TCGGGTCGATGATCTCCGCGAAGCGTCCGGTGTCTGACCCAGGAGCTGTGTGCGGTC 188  
Db 572 TCACGGCGAGGTCTCTCCGCGAGGCGTGAATCTCATGCGCCACGAGGCTTCGCGGATCA 631  
QY 189 TCTGACCGCTCCCTCGTCGATCGCTGCTACTGTAAGCGCCCGACGCTTGGGTTCCAGT 248  
Db 632 TCTCGCTGACGCCCGGCCGGAATCATGTGACGCGAGCACTGTGCTGCGGGGATCGG 691  
QY 249 AGTTGCGGTGGAACCGACCCCTTTGTGCTGCTTACCTGCGCGCCACCTGGAGGAATG 308  
Db 692 AGAGGATCTTATGAAGCTTTGCTCTGCTGTTGATCTTCGCCGGCTGTGGCGGTGA 751  
QY 309 TCGGAACGCTTACGGCAT 326  
Db 752 ACGGGAAGCGTCCGACCT 769

## RESULT 3

US-09-252-991A-10493  
; Sequence 10493, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10493  
; LENGTH: 1260  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10493

Query Match 2.8%; Score 41.2; DB 4; Length 1260;  
Best Local Similarity 50.5%; Pred. No. 0.024; 98; Indels 0; Gaps 0;  
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 129 TCGGGTCGATGATCTCCGCGAAGCGTCCGGTGTCTGACCCAGGAGCTGTGTGCGGTC 188  
Db 114 TCAGCGCGAGGTCTCTCCGCGAGGCGTGAATCTCATGCGCCACGAGGCTTCGCCGATCA 173  
QY 189 TCTGACCGCTCCCTCGTCGATCGCTGCTACTGTAAGCGCCCGACGCTTGGGTTCCAGT 248  
Db 174 TCTCGCTGACGCCCGGCCGCGATCATGTGACGCGGAGCACCTGTGCTGCGGGCATCGG 233  
QY 249 AGTTGCGGTGGAACCGACCCCTTTGTGCTGCTTACCTGCGCGCCACCTGGAGGAATG 308  
Db 234 AGAGGATCTTATGAAGCTTTGCTCTGCTGTTGATCTTCGCCGGCTGTGGCGGTGA 293  
QY 309 TCGGAACGCTTACGGCAT 326  
Db 294 ACGGGAAGCGTCCGACCT 311

RESULT 4  
US-09-252-991A-10567/c  
; Sequence 10567, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10567  
; LENGTH: 1635  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10567

Query Match 2.8%; Score 41.2; DB 4; Length 1635;  
Best Local Similarity 50.5%; Pred. No. 0.027; 98; Indels 0; Gaps 0;  
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 129 TCGGGTCGATGATCTCCGCGAAGCGTCCGGTGTCTGACCCAGGAGCTGTGTGCGGTC 188  
Db 1307 TCAGCGCGAGGTCTCTCCGCGAGGCGCTGAATCTCATGCGCCACGAGGCTTCGCCGATCA 1248  
QY 189 TCTGACCGCTCCCTCGTCGATCGCTGCTACTGTAAGCGCCCGACGCTTGGGTTCCAGT 248  
Db 1247 TCTCGCTGACGCCCGGCCGCGATCATGTGACGCGGAGCACCTGTGCTGCGGGCATCGG 1188  
QY 249 AGTTGCGGTGGAACCGACCCCTTTGTGCTGCTTACCTGCGCGCCACCTGGAGGAATG 308  
Db 1187 AGAGGATCTTATGAAGCTTTGCTCTGCTGTTGATCTTCGCCGGCTGTGGCGGTGA 1128  
QY 309 TCGGAACGCTTACGGCAT 326  
Db 1127 ACGGGAAGCGTCCGACCT 1110

RESULT 5  
US-09-252-991A-10674/c  
; Sequence 10674, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

```
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10674
; LENGTH: 1743
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10674

Query Match      2.8%; Score 41.2; DB 4; Length 1743;
Best Local Similarity 50.5%; Pred. No. 0.028;
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 129 TCGGGTCGATGATCTCCGCGGAACGGTGGGGTGTCTCGACCCAGGAGCTGGTGTGGGTC 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1663 TCAGCGGAGGTCTCTCCGCGGAGGCGCTGAACCTCCATGCCACGCGAGGCTTCGCCGATCA 1604

QY 189 TCTGACCGCTGCCCTCGTGTGATCGGCTGCATCTGAACCGCCCGAGCTTGGGTCCACGT 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1603 TCTCGTGTACGCGCGCGGATCATGTGCACGCGGACCTGTGTGCTGCGGGCATCGG 1544

QY 249 AGTTGCCGTGGAACGACCCCTTCTGCTGCTTACCTGCGCGCCACCTTGGAGGAATG 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1543 AGAGGATCTGTATGAAGCCCTTCCCTCTCGTGGTGTATCTTCGCCCGGCTTGGCGCTGA 1484

QY 309 TCGGAACGCTTACGGCAT 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1483 ACGGAGGCTCGACCT 1466
```

```
RESULT 6
US-09-470-191-27
; Sequence 27, Application US/09470191
; Patent No. 645633
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods of Their Use in
; TITLE OF INVENTION: The Treatment, Prevention and Diagnosis of Tuberculosis
; FILE REFERENCE: 014058-008910US
; CURRENT APPLICATION NUMBER: US/09/470,191
; CURRENT FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 60/113,952
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)...(551)
; OTHER INFORMATION: n = any nucleotide
US-09-470-191-27

Query Match      2.7%; Score 38.8; DB 4; Length 551;
Best Local Similarity 44.7%; Pred. No. 0.083;
Matches 139; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 97 ATGAAGCCGCCACGCTGATGGCTTCAATGCTTCGGGTGCGATGATCTCCCGGAACGGTG 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 ATCCCGCGCGGCGCGGATGGCTCGGCAATGGTTCGGATNGTGGCGGCAACCGANCGTTG 233

QY 157 CGGGTGTCTACCCAGGAGCTGGTGTGCGGTCTCTGACCGCTGCCCTCGTCTGATCGGCTG 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 GGCGTGGCTACCATCTCCGACGCTGACGGCGGTACACCCAGCGCTCGCGAGCGCGG 293

QY 217 CATACTGAACGCCCGACGTTTGGCGGTCCACGTAGTTGCGGTGGAACCGACCCCTTTGTC 276
```

```
Db 294 CGGGGTCCACCGCGCGAACTGACGTCTTCGGCGGCGCGTCCGTCATCGGTGAAGGCATC 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 277 GTGCTTACCTGCGCGCCACCTTGAGGAACTCTCGGAACGCTTACGGCATCTCTACGAC 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 GTCCGTGCGCGGGGACAGCCGGAATTGGCAATCGGAACCCCGGCAATCCGAGCGATCGC 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 337 CTAGTCGAGGGGAGCTATTGTTGTTAAACCAAGGGGCAAGCTGATGGGCCCGTTCCGATCTT 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 CTCGCGGGGTACACACACGCCCAAGCGCGCGGGTTCGGGNTTTCGCGNATTAACCTT 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 397 GTAATTGCTAA 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 GTTGTGCGAA 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 7
US-09-252-991A-15931
; Sequence 15931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15931
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15931

Query Match      2.6%; Score 38; DB 4; Length 507;
Best Local Similarity 48.6%; Pred. No. 0.14;
Matches 104; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 145 CGCGGAACGTCGCGGTGTCTGACCCAGGAGCTGTGTCGGGTCTCTGACCGTGCCTC 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 CACAGCGCTGCGGGTACGTACCGCGCAAGGGCTGTTCAGGCCCTTCGGCACCCGTTTC 119

QY 205 GTGATCGGCTGCATCTGAACGCCCGGAGCTTCCAGTGTGCGGTCCAGTACTTCCGTTGGAACC 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 AACGTACGCTGCAAGATGCCGCCACCTGTCTCAGCGTCAGCGAGGTTTCGGTGGCATC 179

QY 265 GACCCCTTGTGTCGCTTACCTGCGCGCCACCTCGAGGAATGTCCGAACGCTTACGGC 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 GATGCTTCGGCGGGTCCGCTGAGGCTCCGTCGCGGCGGAGGCGGAGAGTATCGG 239

QY 325 ATCTCTACGACCTAGTCGAGGGCGACTATTTCG 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 ATCGCGCGCGCGCGCTGCGGCGCTCGAACGCG 273
```

```
RESULT 8
US-09-252-991A-15895
; Sequence 15895, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
```

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15895  
; LENGTH: 1056  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15895

Query Match 2.6%; Score 38; DB 4; Length 1056;  
Best Local Similarity 48.6%; Pred. No. 0.2;  
Matches 104; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 145 CGCGAAGCGTGGGGTCTCGACCCAGGAGCTGGTGGGTCTCTGACCGCTGCCCTC 204  
Db CACAGGCGCTGGGGTACGCGCAAGGGCTGTCGAGGCCCTCGGCACCGGTTTC 651  
QY 205 GTGATCGCTGCATCTGAAGCCCGACGCTGGGTTCACGTAGTTGGCGTGAAC 264  
Db AACGTACCGCTGCAAGATGTCGCGCCACCTGTCTCAGCGTCAAGAGGTTCTGGTGGCATC 711  
QY 265 GACCCCTTTGTGTCGCTTACTTGGCGCCACCTGGAGGAATGTCGGAACGCTTACGC 324  
Db GATGCTTCGGCGCGCTCCGCTGAGGCTCGGTGGCCGAGGCGGCGAGCTATCGG 771  
QY 325 ATCTCTACGACCTAGTCGAGGGCGACTATTTC 358  
Db ATCGCCGCGAGCGGCTGCGCGCTCTGAACGGC 805

## RESULT 9

US-09-252-991A-15868  
; Sequence 15868, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15868  
; LENGTH: 3357  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15868

Query Match 2.6%; Score 38; DB 4; Length 3357;  
Best Local Similarity 48.6%; Pred. No. 0.36;  
Matches 104; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 145 CGCGAAGCGTGGGGTCTCGACCCAGGAGCTGGTGGGTCTCTGACCGCTGCCCTC 204  
Db CACAGGCGCTGGGGTACGCTACGCGCAAGGCTGTCGAGGCCCTTCGCGACCGGTTTC 346  
QY 205 GTGATCGCTGCATCTGAAGCCCGACGCTGGGTTCACGTCGAGTGGCGTGAAC 264  
Db AACGTACCGCTGCAAGATGTCGCGCCACCTGTCTCAGCGTCAAGAGGTTCTGGTGGCATC 406  
QY 265 GACCCCTTTGTGTCGCTTACTTGGCGCCACCTGGAGGAATGTCGGAACGCTTACGC 324  
Db GATGCTTCGGCGCGCTCCGCTGAGGCTCCGTTGGCCGAGGCGGCGAGCTATCGG 466  
QY 325 ATCTCTACGACCTAGTCGAGGGCGACTATTTC 358  
Db ATCGCCGCGAGCGGCTGCGCGCTCTGAACGGC 500

## RESULT 10

US-09-252-991A-4342/c  
; Sequence 4342, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4342  
; LENGTH: 1257  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4342

Query Match 2.6%; Score 37.8; DB 4; Length 1257;  
Best Local Similarity 50.6%; Pred. No. 0.25;  
Matches 117; Conservative 0; Mismatches 112; Indels 2; Gaps 1;  
QY 149 GAACGGTGGGGTCTCGACCCAGGAGCTGGTGGGTCTCTGACCGCTGCCCTCGTCG 208  
Db GAGCTGATGACATCTGCAAGAGGAGGAGGCGGCGGCGCAAGGTCCTGGGCTACACG 475  
QY 209 ATCGGCTGCATCTGAACCGCCGACCTTGGGTCCACGTAGTTGCGTGGAAACCGACC 268  
Db GTCTATACCGCGCGCGCACACCTCGGCTCGAGGTGTTGC--TGGAGCAGGAAG 417  
QY 269 CTTTGTGTCGCTTACTTGGCGCCACCTTGGAGGAATGTCGGAACGCTTACGSCATCT 328  
Db GCTTCAAGTGGCGGTCTCGCGCGAGCGTGGATGCCAGCGCGCGGAAGACTGGATCG 357  
QY 329 CCTACGACCTAGTCGAGGGCGACTATTGTTCTTAACGAAGGGGCGCAAGCTGG 379  
Db CCGAGCAACTGGACCGTGGCATCGACGTCTCATCAACCAACCGCGAGTTGG 306

## RESULT 11

US-09-252-991A-4513  
; Sequence 4513, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4513  
; LENGTH: 1353  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4513

Query Match 2.6%; Score 37.8; DB 4; Length 1353;  
Best Local Similarity 50.6%; Pred. No. 0.26;  
Matches 117; Conservative 0; Mismatches 112; Indels 2; Gaps 1;  
QY 149 GAACGGTGGGGTCTCGACCCAGGAGCTGGTGGGTCTCTGACCGCTGCCCTCGTCG 208  
Db GAGCTGATGCACATCTGCAAGAGGAGGAGGCGGCGGCGCAAGGTCCTTGGGCTACACG 882  
QY 209 ATCGGCTGCATCTGAACCGCCCGACGTTGGCGTCCACGTAGTTGCGTGGAAACCGACC 268

Db 883 GTCTATACCGGACGCGGACACACCTCGCGCTGAAGGTGTC--TGAGCAGGAAG 940  
QY 269 CTTTGTGCTGCTTACTGCGCGCACCTTGGAGAAATGTCGGAACGCTTACGGCACT 328  
Db 941 GCTTCAAGGTGCGGTGCTGCGCGGAGCGTGGATGCCAGCGCGCGGAAGACTGGATCG 1000  
QY 329 CCTACGACTAGTCAGGCGGACTATTGCTTAAACAAGGGGCCAAGCTGG 379  
Db 1001 CCGAGCAATGGACCGTGCACTGACGCTCATCACCAACCCGAGTTGG 1051

RESULT 12  
US-09-252-991A-15497/c  
; Sequence 15497, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15497  
; LENGTH: 1419  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15497

Query Match 2.6%; Score 37.8; DB 4; Length 1419;  
Best Local Similarity 50.3%; Pred. No. 0.27;  
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 125 TGCCTCGGTGATGATCTCCGCGGAACGCTGCGGTGCTCGACCCAGGAGCTGGTGTG 184  
Db 924 TCCTTCTGTACAGCAGGTGCTGGACAGCGCGGCTGCGCCAGGAAGTCCAGGCC 865

QY 185 GGTCTCTGACCGCTGCTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 244  
Db 864 AGGAGAGCATGGAAACTCGGCTCTGATGCTGACGCGGTATCGCGTTCGGGTAG 805

QY 245 ACCTAGTTCGCGTGGAAACCGACCCCTTTGCTGCTTACTGACCGCGCCACCTCGAGG 304  
Db 804 ACGCCGTTGAAGCGCACGCGGAGGCGTGTCTCGCGAAGCGCGCGCGGCGATGTCGAGG 745

QY 305 AATGT 309  
Db 744 TGAGT 740

RESULT 13  
US-09-252-991A-15291  
; Sequence 15291, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15291  
; LENGTH: 2016

; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15291

Query Match 2.6%; Score 37.8; DB 4; Length 2016;  
Best Local Similarity 50.3%; Pred. No. 0.32; Mismatches 92; Indels 0; Gaps 0;  
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 125 TGCCTCGGTGATGATCTCCGCGGAACGCTGCGGTGCTCGACCCAGGAGCTGGTGTG 184  
Db 1520 TCCTTCTGTACAGCAGGTGCTGGACAGCGCGGCTGCGCCAGGAAGTCCAGGCC 1579

QY 185 GGTCTCTGACCGCTGCTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 244  
Db 1580 AGGAGAGCATGGAAACTCGCGGCTCTGATGCTGACGCGGTATCGCGTTCGGGTAG 1639

QY 245 ACCTAGTTCGCGTGGAAACCGACCCCTTTGCTGCTTACTGCGCGCCACCTCGAGG 304  
Db 1640 ACGCCGTTGAAGCGCACGCGGAGGCGTGTCTCGCGAAGCGCGCGCGGATGTCGAGG 1699

QY 305 AATGT 309  
Db 1700 TGAGT 1704

RESULT 14  
US-09-252-991A-15550/c  
; Sequence 15550, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15550  
; LENGTH: 2508  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15550

Query Match 2.6%; Score 37.8; DB 4; Length 2508;  
Best Local Similarity 50.3%; Pred. No. 0.35; Mismatches 92; Indels 0; Gaps 0;  
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 125 TGCCTCGGTGATGATCTCCGCGGAACGCTGCGGTGCTCGACCCAGGAGCTGGTGTG 184  
Db 1154 TCCTTCTGTACAGCAGGTGCTGGACAGCGCGGCTGCGCCAGGAAGTCCAGGCC 1095

QY 185 GGTCTCTGACCGCTGCTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 244  
Db 1094 AGGAGAGCATGGAAACTCGCGCTCTGATGCTGACGCGGTATCGCGTTCGGGTAG 1035

QY 245 ACCTAGTTCGCGTGGAAACCGACCCCTTTGCTGCTTACTGCGCGCCACCTCGAGG 304  
Db 1034 ACGCCGTTGAAGCGCACGCGGAGGCGTGTCTCGCGAAGCGCGCGCGGATGTCGAGG 975

QY 305 AATGT 309  
Db 974 TGAGT 970

RESULT 15  
US-09-252-991A-4633/c  
; Sequence 4633, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2004, 03:13:49 ; Search time 757.822 Seconds  
(without alignments)  
9773.579 Million cell updates/sec

Title: US-10-668-047-1  
Perfect score: 1461  
Sequence: 1 gtgcatcgatcgccagcac.....ttccggaatcgctcgta 1461

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.8	3.1	536	17	US-10-338-110-119
2	44.2	3.0	927	17	US-10-437-963-31196
3	44	3.0	3316	16	US-10-369-493-41493
4	43.8	3.0	1966	16	US-10-369-493-31580
5	42.8	2.9	2889	17	US-10-437-963-55282
6	42.2	2.9	1947	16	US-10-381-779-10
7	41.2	2.8	780	15	US-10-156-761-6003
8	41.2	2.8	1425	15	US-10-282-122A-14027
9	41.2	2.8	9025608	15	US-10-156-761-1
10	41	2.8	1311	13	US-10-282-122A-14145
11	40.6	2.8	2871	16	US-10-369-493-47097
12	39.8	2.7	1401	15	US-10-156-761-6097
13	39.8	2.7	9025608	15	US-10-156-761-1
14	39.6	2.7	1503	13	US-10-282-122A-13659

C 15	39.4	2.7	987	13	US-10-282-122A-31696	Sequence 31696, A
C 16	39.4	2.7	1119	13	US-10-282-122A-11706	Sequence 11706, A
C 17	39	2.7	975	17	US-10-437-963-32538	Sequence 92538, A
C 18	39	2.7	1287	15	US-10-156-761-6510	Sequence 6510, Ap
C 19	39	2.7	13494	15	US-10-156-761-2355	Sequence 2355, Ap
C 20	38.8	2.7	2541	15	US-10-156-761-2855	Sequence 2855, Ap
C 21	38.8	2.7	9810	10	US-09-373-658-35	Sequence 35, Appl
C 22	38.8	2.7	9810	11	US-09-989-687-35	Sequence 35, Appl
C 23	38.2	2.6	597	17	US-10-437-963-18193	Sequence 18193, A
C 24	38	2.6	1051	17	US-10-767-701-13578	Sequence 13578, A
C 25	38	2.6	1135	13	US-10-425-114-20794	Sequence 20794, A
C 26	38	2.6	1584	16	US-10-369-493-42282	Sequence 42282, A
C 27	37.6	2.6	1044	15	US-10-156-761-2398	Sequence 2398, Ap
C 28	37.6	2.6	1272	16	US-10-369-493-44349	Sequence 44349, A
C 29	37.6	2.6	1338	15	US-10-156-761-2638	Sequence 2638, Ap
C 30	37.4	2.6	1332	17	US-10-437-963-58487	Sequence 58487, A
C 31	37.2	2.5	549	17	US-10-437-963-57513	Sequence 57513, A
C 32	37.2	2.5	687	13	US-10-282-122A-26299	Sequence 26299, A
C 33	37.2	2.5	690	13	US-10-282-122A-28509	Sequence 28509, A
C 34	37.2	2.5	1341	15	US-10-156-761-1069	Sequence 1069, Ap
C 35	37	2.5	1188	15	US-10-156-761-5738	Sequence 5738, Ap
C 36	37	2.5	1316	17	US-10-437-963-98904	Sequence 98904, A
C 37	36.8	2.5	2029	17	US-10-437-963-91052	Sequence 91052, A
C 38	36.6	2.5	1049	13	US-10-142-426-358	Sequence 358, App
C 39	36.6	2.5	1049	15	US-10-123-155-358	Sequence 358, App
C 40	36.6	2.5	1049	15	US-10-146-731-358	Sequence 358, App
C 41	36.6	2.5	1049	15	US-10-140-472-358	Sequence 358, App
C 42	36.6	2.5	1049	15	US-10-141-761-358	Sequence 358, App
C 43	36.6	2.5	1049	15	US-10-142-885-358	Sequence 358, App
C 44	36.6	2.5	1049	15	US-10-158-790-358	Sequence 358, App
C 45	36.6	2.5	1049	16	US-10-137-871-358	Sequence 358, App

## ALIGNMENTS

## RESULT 1

US-10-338-110-119  
; Sequence 119, Application US/10338110  
; Publication No. US20040023254A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuhrmann, Jeffrey J.  
; APPLICANT: Ronsesser, James A.  
; TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial  
; TITLE OF INVENTION: Communities  
; FILE REFERENCE: HER-0056  
; CURRENT APPLICATION NUMBER: US/10/338,110  
; CURRENT FILING DATE: 2003-01-07  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 119  
; LENGTH: 536  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus Sequence  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (7)..(7)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (21)..(21)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (69)..(69)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:

```
NAME/KEY: misc feature
LOCATION: (87)..(97)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (213)..(213)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (226)..(226)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (261)..(261)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (273)..(273)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (347)..(347)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (357)..(357)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (398)..(398)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (419)..(419)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (509)..(509)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (517)..(517)
OTHER INFORMATION: n is a, c, g, or t
US-10-338-110-119
```

```
Query Match 3.1%; Score 44.8; DB 17; Length 536;
Best Local Similarity 14.0%; Pred. No. 0.00051;
Matches 43; Conservative 151; Mismatches 114; Indels 0; Gaps 0;

QY 1 GTGCATCCGATCCGACGACTGAAGCTCGCGCAAGCTGCTCTCGCAACTGGACCCC 60
DB 228 SBKCRVSRVVRVMYRCSSVSVATCCVMSRTNTGGARHTKWCNGMTWCRSMCR 287

QY 61 ACTACTCAAGCGTGTAGGCGAGTTCTTCACTCCCATGAAGCGCGCCAGCTGATGGCT 120
DB 288 CRHSCSSAMYKGCSTRTGMVKWRTYTSKSDVMRBCSTSVARSMGSCDSYVWSBGGN 347

QY 121 TCAATGCTTCGGTTCGATGATCTCCGCGGAACGGTGGCGGTGCTCGACCCAGAGCTGCT 180
DB 348 GCVRKEVSNTRTSRGSVRSYRSVHCRBSRSATGGWCGSYWKTWCVDNMKSAWCGGB 407

QY 181 GTCCGGTCTCTGACCGCTGCCTGCTGATCGGCTGCATCTACTGAACCCCGCGAGTTCG 240
DB 408 RTWDBSHSANCVSIVGGBYCGSYEMHSHVHVRBRVRDMHRTVSTSGCSTSYG 467

QY 241 GTCCACGTACTGCGTGGAAACGACCCCTTTGCTGCTTACCTGCGCGCCACCTG 300
DB 468 GAYCRVBADCRMDSCVTVRMSVHBYKYVSKYRSSMHYCNRRHBBGVBNVYKYVGV 527

QY 301 GAGGAATG 308
DB 528 DRBYKVTG 535
```

```
RESULT 2
US-10-437-963-31196/c
Sequence 31196, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
NUMBER OF SEQ ID NOS: 2003-05-14
SEQ ID NO 31196
LENGTH: 927
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_35525C.1
US-10-437-963-31196
```

```
Query Match 3.0%; Score 44.2; DB 17; Length 927;
Best Local Similarity 54.7%; Pred. No. 0.0011;
Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 136 GATGATCTCCGCGAAGCGTTCGGGTGCTCGACCCAGGAGCTGTTGTCGGGTCTCTGACC 195
DB 834 GAGGACGCGCGCTTCAGGAGGGTGTGAAGACCCAGGCGCGCATGCCGCTGCACCCGCC 775

QY 196 GTGTCCTCGTCGATCGGCTGCATGTAACGCGCCGAGTTCGGTTCACGTTAGTTGCC 255
DB 774 GCCGCGACGCGCGCGAAGTGCATCCACACGCGCGCGCTGGCAGAGGTTGCACCC 715

QY 256 GTGGAACCGACCCCTTTGTCGTGCTTACCTGCGCGCCAC 296
DB 714 GAGGACGCGCACTGCGACGCGCGAGCGCGGGGGGCCAC 674
```

```
RESULT 3
US-10-369-493-41493/c
Sequence 41493, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 41493
LENGTH: 3316
TYPE: DNA
ORGANISM: SPHINGOMONAS
US-10-369-493-41493
```

```
Query Match 3.0%; Score 44; DB 16; Length 3316;
Best Local Similarity 47.2%; Pred. No. 0.0027;
Matches 134; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
```

Qy	129	TCGGGT	CGATGATCTCCGCGGAACGGTGC	CGGGTGTGCA	CCAGAGATGGTGT	CGGGTC	188
Db	2531	TCTCGGCGATGATCTCGCGGGCAGGCTGTAGGGCGGGATCGACGAGCGGTGTGCGCCG	2472				
Qy	189	TCTGACCGCTGCCCTCGATCGGCTGCATCTACTGAAACGCCGCCACAGTTGCGGTCACAGT	248				
Db	2471	AATGGACGCCGGCTTCTCGATATGCTGGAGCACGCCGCCACAGACGATCTCGGTCCCGT	2412				
Qy	249	AGTTGCGGTGGAAACCGACCGCTTTGTGTCGCTTACTCTCGCGCCACCTTGGAGGAATG	308				
Db	2411	CGGCGATCGGCTCCACGTCCACTCCGATCGCGTTCGCGCAGATATCTGGTCGATCAGCACCG	2352				
Qy	309	TCGGAAACGCTTACGGCATCTCCTACGACCTAGTCTGAGGGGACGATTTGCTTAACCAAGG	368				
Db	2351	GCGAGTCGCCCGAGACCTGCACCGGCTCTGGATATAGTCTGAGCTCTGGAGGCTGT	2292				
Qy	369	GGCCAGCTGGATGGCCCGCTTCGATCTTGTAATGCTAACTCTC	412				
Db	2291	CGACGATCTCCATCGCCCGCGCCGAGCACATAGCTTCGCGCCG	2248				

```

RESULT 4
US-10-369-493-31580
; Sequence 31580, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31580
; LENGTH: 1966
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31580

```

Query Match	3.0%;	Score 43.8;	DB 16;	Length 1966;
Best Local Similarity	49.0%;	Pred. No. 0.0023;		
Matches 117;	Conservative 0;	Mismatches 122;	Indels 0;	Gaps 0;
QY	142	CTCCGCGGAACGTCGGGGTCTCGACCCACGAGAGCTGGTCGCGTCTCTGACCGCTGCC	201	
Db	1313	CGCTTCGTGATCGACCGGCGGGGCTGTGGGGCCGATGGCGGCCCATTCGGGGGGCC		
QY	202	CTCGTCGATCGGTCGCATCTGAAGCCGCCGAGCTTCGCGTCCACGTPAGTTGCCGTGGAA	261	
Db	1373	TTCCAGCGTTGGCTTCATCACTTCGCTGCCCAACATGACCCGTGATGGCGCGGCCGACGAG	1432	
QY	262	ACCGACCCCTTTGTCGTGCCTTACCTGCGCGCCACCTCGAGAGAACTCGGAAGCTTAC	321	
Db	1433	GCCGAGCTCATCCACATGATCGCCACCGCGTGGCCTTCGACGAGGCCCATCGCCCTTC	1492	
QY	322	GGGATCTCCTACGACTAGTCGAGGGGCGACTATTGCTTAAACAGGGGCCAAGCTGGA	380	
Db	1493	CGCTTCCCGCGGAGCGAGGGGTGGCGTTCGAGATGCCGAGGCGCGGACGGTCTGGA	1551	

RESULT 5  
US-10-437-963-55282  
; Sequence 55282, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.

```

? APPLICANT: Zhou, Yihua
? APPLICANT: Cao, Yongwei
? APPLICANT: Wu, Wei
? APPLICANT: Boukharov, Andrey A.
? APPLICANT: Barbazuk, Brad
? APPLICANT: Li, Ping
? TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
? TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
? FILE REFERENCE: 38-21(53221)B
? CURRENT APPLICATION NUMBER: US/10/437,963
? CURRENT FILING DATE: 2003-05-14
? NUMBER OF SEQ ID NOS: 204966
? SEQ ID NO 55282
? LENGTH: 2889
? TYPE: DNA
? ORGANISM: Oryza sativa
? FEATURE:
? OTHER INFORMATION: Clone ID: PAT_MRT4530_57303C.1
? US-10-437-963-55282

```

Query Match	2.9%;	Score 42.8;	DB 17;	Length 2889;
Best Local Similarity	46.1%;	Pred. No. 0.0064;		
Matches 143;	Conservative 0;	Mismatches 167;	Indels 0;	Gaps 0;

  

QY	34	CAAGCTGCTCTCGGCAAACTGGACCCACACTACTCAAGCGGTGCTAGGCGAGTTCTTCTACT	93
Db	286	CAGAGAGTTGCCCAACCTCTCGAGCGCGCTCTCTCGCGGTGGGCGGTTCTGCCTCGGC	345
QY	94	CCCATGAAGCGCGCACGCTGATGGCTTCAATGCTTGGGGTCGATGATCTCGCGGGAACG	153
Db	346	CCGCGCGGTGCTCAGCAGATCTCCCTGGCGTTCGCGGCGCTGGCGCTCTCGGCCGAG	405
QY	154	GTGCGGGTGTGCACCCAGAGCTGGTGTGCGGTCTCTGACCGTGCCTCTGTCGATCGG	213
Db	406	GGCGGGGTGGACGCGCTCTTTCGCGCGATGCGGCACCTCCCGACCCCGCGCTGATGAG	465
QY	214	CTGCATACTGAACGCCCGGACGTTGCGGTCCACGAGTGTGCGGTGGAAACCGACCCCTTT	273
Db	466	CTGCTACTGTGCTCCCGGAGGAGGTGTCGAGGACCCAGACCGGGGTCGAC	525
QY	274	GTGCGTCCCTTACCTGCGCGCCACCTCGAGGAATGTCGGAAACGCTTACGGCATCTCCCTAC	333
Db	526	GC CGCGGGCGCTGCAGGTTTCA CCGCGAGCTCATGGCGACGCGCCGCGCTGCTCGAG	585
QY	334	GACCTAGTGG	343
Db	586	TTCTCTCTCG	595

```

RESULT 6
US-10-381-779-10
; Sequence 10, Application US/10381779
; Publication No US20030219798A1
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-381-779-10

```

Query Match 2.8%; Score 42.2; DB 16; Length 1947;  
 Best Local Similarity 48.5%; Pred. No. 0.0082; Mismatches 123; Indels 0; Gaps 0;  
 Matches 116; Conservative 0;

QY 142 CTCGCCGGAACGGTCCGGGTGCTCGACCCAGGAGTGGTGTGGGTCTCTGACCGCTGCC 201  
 DB 1255 CGCTTCGTGATGACCGGGGGGGTCTGTGGGGCCGATGGCGACCATCGGGGGCC 1314  
 QY 202 CTGTCGATCGCTGATCTGAAACGCGCCGAGCGTGTGGGTCCAGTAGTTCCTGGAA 261  
 DB 1315 TTGACGCTTGGCTTATCATCTTCGTGCTCCCAACATGACCGTGTGGCGCGCCGACGAG 1374  
 QY 262 ACCGACCCCTTGTGCTGCTTACCTGTGGCGCCACCTGGAGGAATGTGGAACGCTTAC 321  
 DB 1375 GCCGAGCTCATCCATGATGCCACCGCGGTGGCTTCGGGAGGGCCCATCGCTTC 1434  
 QY 322 GGCATCTCTACGACCTAGTCTGAGGGCGACTATTTGCTTAAACAAAGGGCCCAAGTGA 380  
 DB 1435 CGCTTCCCGGGGGGAGGGGTGGGCTCGAGATGCCCGAGCGCGGACGGTGTGGA 1493

RESULT 7

US-10-156-761-6003  
 ; Sequence 6003, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 6003  
 ; LENGTH: 780  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(780)  
 US-10-156-761-6003

Query Match 2.8%; Score 41.2; DB 15; Length 780;  
 Best Local Similarity 47.3%; Pred. No. 0.011; Mismatches 138; Indels 0; Gaps 0;  
 Matches 124; Conservative 0;

QY 130 CGGGTGTGATCTTCGCGGAACGGTCCGGGTGTGACCCAGGAGTGTGTGCGGTCT 189  
 DB 85 CGTGGCGCGCGCGCGGGGATGTTGTGCGCGCGGTTCGCGGACCTGTGTGGGTG 144  
 QY 190 CTGACCGCTGCCCTCTGCTGATCGGCTGCATCTGAAACCGCCCGACGTTGCGGTCCAGTA 249  
 DB 145 CTGCGCGCGGCTTCGCGACGCTGCTCTCTGCTGGCGCGGAGCGGTGCTCGCGCC 204  
 QY 250 GTTGGCGTGAACCGACCCCTTTGTGCTGCTTACTGTGGCGGCCACCTGGAGGATGT 309  
 DB 205 GTGGCCACCGCGCGTGGCGCGCTCTCACCGCGGGTCCACTCTGACGCGGTGCC 264  
 QY 310 CGGACGCTTACGCGCATCTCTACGACTAGTCTGAGGCGCACTATTTGCTTAAACGAGG 369  
 DB 265 GACACCGGAGCGCTTCGGGAGCGGGAAGCCCGCGGAGCGGCTGCGGATCATGAG 324  
 QY 370 GCCAAGCTGGATGCCCGCTTCG 391

DB 325 CAGTCGGACATCGGCCGCTTCG 346  
 RESULT 8  
 US-10-282-122A-14027/C  
 ; Sequence 14027, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 14027  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Burkholderia mallei  
 US-10-282-122A-14027

Query Match 2.8%; Score 41.2; DB 13; Length 1425;  
 Best Local Similarity 52.3%; Pred. No. 0.015; Mismatches 83; Indels 0; Gaps 0;  
 Matches 91; Conservative 0;

QY 125 TGCTTCGGTTCGATATCTCCGGGAACGGTCCGGGTGTGACCCAGGAGCTGTGTCG 184  
 DB 236 TGGCGCGCTGATACGCTGCTGTGTGTGTGACGCTGACGACCGCGCGCG 177  
 QY 185 GGTCTCTGACCGCTGCCCTCTGATCGGTGGTGCATACCTGAACCGCCCGACGTTGCGGTCC 244  
 DB 176 AGCGCGCGCGCTTGCCTCGACCACTGCTCCGCAAGCCAGTTCGTTCAGATTGCCGATC 117  
 QY 245 ACCTAGTTCGCTGGAACCGACCCCTTTCTGCTGCTTACCTGCGCGCCACCC 298  
 DB 116 GCGACGATCGGTGCGCGGTATCGCGACGCTCGGCTCGACGCTCGCGCGCGCC 63

RESULT 9

US-10-156-761-1/c  
 ; Sequence 1, Application US/10156761

```
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      2.8%; Score 41.2; DB 15; Length 9025608;
Best Local Similarity 47.3%; Pred. No. 2.3; Mismatches 138; Indels 0; Gaps 0;
Matches 124; Conservative 0;

Qy 130 CGGTCGATGATCTCCGGGAACGTCGGGTGCTCGACCCAGGAGCTGTGTCGGGTCT 189
Db 7256910 CGTCCGCGCGCGCGCGGGATGTGTGCGCGCGGTGCGGACTGTGTCGGGTG 7256851

Qy 190 CTGACCGCTGCCCTCGTCGATCGGTCGATCTGAACGCCCGGACGTTGCGGTCCACGTA 249
Db 7256850 CTCGCCGCGGCGCTCGGCACGCTGTCTCTGTCGGCGGAGCCCGTTGCTCGCGCC 7256791

Qy 250 GTTCCGCTGGAACCGACCCCTTTGTCGTGCTTACCTGCGCGGCCACCTTGAACCAAGG 309
Db 7256790 GTGGCCACCGCGCGCGCGCTGCTCTCTGTCGGCGGAGCCCGTTGCTCGCGCC 7256731

Qy 310 CGGAACGCTTACGGCATCTCTACGACCTAGTCGAGGCGGACTATTGCTTAAACCAAGG 369
Db 7256730 GACACCGGAGCGGCTCGGCGAGCGCAAGCCCGCGAGGACGCGTCCGATCATGAG 7256671

Qy 370 GCCNAGCTGATGCCCGCTTCG 391
Db 7256670 CAGTCGACATCGGCGCGCTTCG 7256649

RESULT 10
US-10-282-122A-14145
Sequence 14145, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14145
LENGTH: 1311
TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-14145

Query Match      2.8%; Score 41; DB 13; Length 1311;
Best Local Similarity 46.9%; Pred. No. 0.017; Mismatches 128; Conservative 0; Indels 0; Gaps 0;
Matches 128; Conservative 0;

Qy 108 CACGCTGATGCTTCAATGCTTCGGGTGCGATCTCCGCGAAGCGTTCGGGTGCTCGA 167
Db 696 CACGCGCTCGCGCGGATGCGGTGCGCGGAGATGATGACGACGCGCGGTGCGGA 755

Qy 168 CCCAGGAGCTGCTGCGGTCTCTGACGCTGCCCTGTCGATCGGTGCATATGTAACG 227
Db 756 CGGCGCTGTCGCGGACGCTCGATTGATGACGACATGTCGACCATCTCTCGTGTTCG 815

Qy 228 CCCGAGCTTGCCTCCAGCTAGTTCGCGTGGAAACGACCCCTTTGTCGTCCTTACCT 287
Db 816 GCACGGGACGCGGACGCGAGCGGCGGTCGCGGTGACCGCGTGCAGCGGATCGC 875

Qy 288 GCGCGCCACCTGGAGGAATGTGGAACGCTTACGGCATCTCTTACGACCTAGTCGAGGG 347
Db 876 GCGCACTTACGCGCGGTGTCCCGAAGCGGCCACCGTGCACCGGCTCGGCGCGG 935

Qy 348 CGACTATTGCTTAAACCAAGGGGCAAGCTGGA 380
Db 936 GCGGGGCTTTGCGCTGCGCGGCGGCGAGCTCGA 968

RESULT 11
US-10-369-493-47097/c
Sequence 47097, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
```





GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 13659

LENGTH: 1503

TYPE: DNA

ORGANISM: Burkholderia fungorum

US-10-282-122A-13659

Query Match 2.7%; Score 39.6; DB 13; Length 1503;

Best Local Similarity 50.8%; Pred. No. 0.055;

Matches 120; Conservative 0; Mismatches 114; Indels 2; Gaps 1;

Qy	138	TGATCTCCGCGAACGCTGCGGGTCTCGACCCAGGAGCTGGTCTCGGTTCTGTGACCGC	197
Db	431	TCATTGCGGACGCGGTCTCACGACTCCACCCGAGCTGCTCACCGGCGCTTCGGCC	490
Qy	198	TGCCTCTCGATCGGTGTCATCTGAACGCCCGACGTTGCGGTCCACGTTAGTTCGCGT	257
Db	491	TTACCGCGCGCGGACGCGGTGTGAGCGCGGAGCTCGCTGTATCCGACCTGTCT	550
Qy	258	GGAACCGACCCCTTTCTGCTTCTACCTGCGGCGCACCGCTGAGGAATGTCGGAACGC	317
Db	551	GGCTGCGGTGCGCGACGCGCGGAAGCGCGGCTGAAACGCGCGATCACCGTGATGA	610
Qy	318	TTACGGCATCTCTTACACCTAGTCGAGGCGGACTATTTCCTTAACCAAGGGCCA	373
Db	611	CCACGNAACCGTCATGAC--AGTCGACCGGACACCCCTGCTGACGACGACGCA	664

RESULT 15

US-10-282-122A-31696/c

Sequence 31696, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 31696

LENGTH: 987

TYPE: DNA

ORGANISM: Pseudomonas putida

US-10-282-122A-31696

Query Match 2.7%; Score 39.4; DB 13; Length 987;

Best Local Similarity 54.5%; Pred. No. 0.051;

Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy	134	TCGATGATCTCCGCGAACGCTGCGGGTCTCGACCCAGGAGCTGGTGTGCGGTTCTGA	193
Db	584	TTGCTGAATCGACGAGCGCGGCGCTGTGTCGACGAGCTGTGTCGCGCAGGCTCG	525
Qy	194	CCGCTGCCCTTCGTGATCGGCTGCATCTGAACGCCCGGACGTTGCGGTCCACGTAGTTG	253
Db	524	GTGCTGTGTGATCCAGTCCAGTCCAGCATGTTGGTGTTCGCTGTGTTGACCGCCACTTG	465
Qy	254	CCGTGGAACCGACCCCTTTGTCTGT	278
Db	464	CCGTGCGGGCTGACCGCATGCCTT	440

Search completed: October 2, 2004, 08:33:22

Job time : 780.822 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 16.8208 Seconds  
(without alignments)  
2779.248 Million cell updates/sec

Title: US-10-668-047-2  
Perfect score: 2469  
Sequence: 1 VPIASTETRRQAALGKIDP.....QETLDVAVVRLFCIPESAS 486

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	667.5	27.0	507	1 XYORS	site-specific DNA-
2	662.5	26.8	501	1 S35515	site-specific DNA-
3	379	15.4	546	2 AF2268	type II site-speci
4	243.5	9.9	236	2 AE2268	site-specific DNA-
5	184	7.5	421	2 JN0257	site-specific DNA-
6	180.5	7.3	428	2 JH0634	site-specific DNA-
7	169.5	6.9	377	1 S01615	site-specific DNA-
8	165	6.7	487	2 G64577	type I restriction
9	163	6.6	540	1 J00470	site-specific DNA-
10	162.5	6.6	531	1 XPSTVA	site-specific DNA-
11	161	6.5	543	2 H71935	type I restriction
12	151	6.1	561	2 JN0797	site-specific DNA-
13	148	6.0	579	2 I40371	methyltransferase
14	147	6.0	379	1 P64633	site-specific DNA-
15	141	5.7	571	2 H97333	site-specific modi
16	141	5.7	750	2 AH3158	hypothetical prote
17	141	5.7	832	2 F98128	insertion element
18	140	5.7	555	2 E87003	probable DNA methy
19	137	5.5	514	2 F81152	type I restriction
20	135	5.5	553	2 H70977	hypothetical prote
21	134	5.4	514	2 JH1952	type I site-specif
22	134	5.4	580	2 JH0224	site-specific meth
23	133.5	5.4	381	2 A71882	type II DNA modifi
24	132	5.3	1181	2 B64516	hypothetical prote
25	131.5	5.3	620	2 JC4925	XmI methyltransfe
26	130	5.3	520	2 S02166	type I site-specif
27	130	5.3	545	1 F64579	site-specific DNA-
28	128	5.2	518	1 F64073	site-specific DNA-
29	126	5.1	545	2 F71932	type II DNA modifi

30 125.5 5.1 317 2 E70603 probable dimethyla  
31 122.5 5.0 1252 2 D71810 probable type II D  
32 121.5 4.9 531 2 T30323 type I site-specif  
33 120 4.9 348 2 E82933 type I restriction  
34 119 4.8 823 2 F64526 adenine/cytosine D  
35 118.5 4.8 1261 2 A13369 vitamin B12-depend  
36 118 4.8 1372 2 F97722 hypothetical prote  
37 117 4.7 294 2 B86711 kasugamycin dimeth  
38 117 4.7 518 2 A89808 hypothetical prote  
39 116.5 4.7 279 2 C95232 dimethyladenosine  
40 116.5 4.7 285 2 F65035 hypothetical prote  
41 116.5 4.7 289 2 A98217 hemk protein homol  
42 116.5 4.7 289 2 A13069 protoporphyrioxogen  
43 116 4.7 317 2 E95239 conserved hypothet  
44 116 4.7 317 2 E98103 conserved hypothet  
45 116 4.7 815 2 C71810 type I restriction

ALIGNMENTS

RESULT 1  
XYORS  
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) PstI - Providencia ;  
N:Alternate names: modification methylase PstI; restriction-modification system PstI  
C:Species: Providencia stuartii  
C>Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 05-May-2000  
C:Accession: A00553  
R:Walder, R.Y.; Walder, J.A.; Donelson, J.E.  
J. Biol. Chem. 259, 8015-8026, 1984  
A:Title: The organization and complete nucleotide sequence of the PstI restriction-modif:  
A:Reference number: A92485; MUID:84239756; PMID:6330092  
A:Accession: A00553  
A:Molecule type: DNA  
A:Residues: 1-507 <WAL>  
A:Cross-references: GB:K02081; NID:g150922; PIDN:AAA25672.1; PID:g455323  
A:Experimental source: strain 164  
A:Note: the authors translated the codon CAA for residue 456 as Glu  
C:Comment: Providencia, a genus of enterobacteriacean rods, is sometimes included in the  
C:Function: <Met>  
A:Description: This site-specific methylase recognizes the double-stranded hexanucleotid  
y PstI endonuclease (see PIR:NDOFS).  
C:Superfamily: site-specific DNA-methyltransferase (adenine-specific) PstI  
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 27.0%; Score 667.5; DB 1; Length 507;  
Best Local Similarity 34.2%; Pred. No. 2.2e-43;  
Matches 166; Conservative 87; Mismatches 206; Indels 27; Gaps 10;

Qy 11 RQALGKLDPTQAVLGQFTTPKKAATLMASMLRVDDLGRVTVLDGAGVGSITAAIYD 70  
Db 20 REAANSTLDLTRSKUGQFMSSSAVSELMANLF--ESYVGEHEILDAGAGVGSUTAAFPV 77  
Qy 71 RLHTRPDVAVHVVAVETDPFVPLRATLEECR---NAVGISYD--LVEGDYLLNQAK 125  
Db 78 NA-TLNGAKSISSCYSEI SEVMVYNNLIQVLDLCKIRAMEFEVNWQKIIESDPTQASVEQ 136  
Qy 126 L-----DGPDLVIANPPYKGLASDSLARLATTARAVDVPNVYVAVFVRAVISLKEQGRG 180  
Db 137 LLTIENSPKYNKATLNPYPYLKIAKGRERALLQKVGIEASNLVSFAVALAKQLKSGGEL 196  
Qy 181 VFTVPSRWANGPYRQPRHWMATVSLDILHVESRTRKVFADTKVKQENIVAFSVRPS 240  
Db 197 VAITPRSCFNPDPFRKQMLDECSLNKTHVFNRSKSAFKADNVLOLTIYHLTKGETQ 256  
Qy 241 SSV--VLSRSVAHGEESIASVPPSALVHDEDDDKIVHFA-----ESASVPSAARFLA 292  
Db 257 RKVVTVYSSTCANDINPTIEFVDFEIVKSNNDPLFIHIVTNEQERLANAGGLPCSL 316  
Qy 293 DLGIGVSTGKVDFRNRYQLTDNLDSAGVYPMVYQSNIRSGKIDWPQVGARKPGGFVAVE 352  
Db 317 DLGIGVSTGKVDFRTRNLSMEY-ISNSVPLIFPQHLQRCISIVWPITKAKPNALI-VN 374



```
QY 6 STETRRQAALGKLDPTTQAVLGGQFTPMKAAATLMSMLRVDLDRGTVRVLDPGAGVGSIT 65
Db 8 STDNIRNFSSRLNLHRRRELQGLTTPAPLAFMVG--QFNSLLGHISLIDEGAGVGSIT 65
QY 66 AALVDRLHTRPDV-AHVHVAVEDTPFVVPVLRATLECRNAY---GI--SYDLVEGDYL 119
Db 66 AAFVERLLANSHEVSKCFTVVEPIFILPKQCLDCCCTALENKGAANYCYEKNF- 124
QY 120 LNOGAKLDGP-----FDLVIANPPYKGLASDSLARLATTARAVDVPNVVAFVVRVAV 171
Db 125 IDANSEINPLSTLSAINFTHAIINPLKYKNNKSIEKKIISQMGIEIVNLYSAPVWLT 184
QY 172 ISLKEQGRGVFIVPRSWANGPYRQFRHMLMTAVSLD 208
Db 185 LRLVDEGEIVAITPRFCNGSYFRHFRKSFLEQMKLE 221
RESULT 5
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) TaqI - Thermus aqu
C/Species: Thermus aquaticus
C/Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 05-May-2000
C/Accession: JN0257; S06298
R:Barany, F.; Slatko, B.; Danzitz, M.; Cowburn, D.; Schildkraut, I.; Wilson, G.G.
Gene 112, 91-95, 1992
A/Title: The corrected nucleotide sequences of the TaqI restriction and modification enz
A/Reference number: JN0257; MUID:92201707; PMID:1551602
A/Accession: JN0257
A/Molecule type: DNA
A/Residues: 1-421 <BAR>
A/Note: this is a revision to the sequence from S06298
R:Slatko, B.E.; Benner, J.S.; Jager-Quinton, T.; Moran, L.S.; Simcox, T.G.; Van Cott, E.
Nucleic Acids Res. 15, 9781-9796, 1987
A/Title: Cloning, sequencing and expression of the Tag I restriction-modification system
A/Reference number: S06298; MUID:98096538; PMID:2827113
A/Accession: S06298
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-12, 'A', 14-325, 'Y', 327-351, 'GGRSSTSC', 360, 'RKV', <SLA>
A/Cross-references: GB:Y00499; NID:9288593; PIDN:CAA66551.1; PID:G288594
A/Note: it is uncertain whether Met-1 or Met-31 is the initiator
C/Genetics:
A/Genes: tagIM
C/Function:
A/Description: this enzyme is a methyltransferase of TaqI restriction-modification syste
A/Note: recognition sequence of this enzyme is 5'-TCGA
C/Keywords: methyltransferase; S-adenosylmethionine
Query Match 7.5%; Score 184; DB 2; Length 421;
Best Local Similarity 23.5%; Pred. No. 2.5e-06;
Matches 109; Conservative 58; Mismatches 177; Indels 120; Gaps 20;
QY 26 LGQFFTPMKAATLMSMLRVDLDRGTVRVLDPGAGVGSITAAALVDRLHTRPDVAVHVA 85
Db 18 LGVETPPVDPVFWSLAEP--RGG-RVLEPCAHGPF--RAFREAHGTAHYFVG 69
QY 86 VETDPEW---PYLRATLEECRNAYGISDLVREGDYLLNQAGKLDGPFOLVIANPPYG-- 140
Db 70 VEIDPKALDLPWAEGIL-----ADFLLWEPGE---AFDILGNPPYGIV 111
QY 141 -----KLASDSLARLATTARAVDVPNVVAFVVRVAVISLKEQGRGVFIVPRSWA 189
Db 112 GEASKYPIHVFAVKDLYKKAFTWKG--KYNLYGAFLEKAVRLLKPGGVLPVFPVPAWL 169
QY 190 NGPYRQFRHMLMTAVSLDILHVFESRTKVFDATKVKQENVIVAFSVRQSSSVLRSV 249
Db 170 VLEDFALLREFLAR-----EGKTSVYILGEVFPQKVSADVIRFPKSGKGLSLWD 219
QY 250 AHGEESIASVPFSALVHDEDDDKIVHF--AESASVPSAARFTLADLIGIGVSTGKVVDREN 308
Db 220 TQSESGFTPILWAEYPHWEGE--IIRFETEETRKLEISGMPLGDLFHIRFAARSPFKK 277
QY 309 RQYLTNLDASGVVPMVYQNSIRSGKIDW-----PQVGARKPQGFVAVEDVALRQL 359
```

```
Db 278 HPAVRKE--PGSLFVPLTGRNLKKGWDYKHNHSLWMPKERAKELRDFHAT----- 328
QY 360 LPQGSYVVVKQTKAKEDRRRRVIAAVWDGAS-----RVALDNKTNLYHESQRP 406
Db 329 ----PHLVV----AHTKGRVVA--WDERAYPWREEPHLLPKEGVRLDPST----- 370
QY 407 LEKNVARGMLMLNSTVLDQY----FRAFSGHTQVNGDLRLP 446
Db 371 -----LVQWLNSEAMQKHVRTLYRDFVPHLTRM--LERLP 404
RESULT 6
JH0634
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) TthHB81 - Thermus a
C/Species: Thermus aquaticus
C/Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 05-May-2000
C/Accession: JH0634
R:Barany, F.; Danzitz, M.; Zebala, J.; Mayer, A.
Gene 112, 3-12, 1992
A/Title: Cloning and sequencing of genes encoding the TthHB81 restriction and modificati
A/Reference number: JH0634; MUID:92201699; PMID:1339363
A/Accession: JH0634
A/Molecule type: DNA
A/Residues: 1-428 <BAR>
A/Note: the source was designated as Thermus thermophilus HB8
C/Genetics:
A/Genes: tthHB81M
C/Keywords: methyltransferase; S-adenosylmethionine
Query Match 7.3%; Score 180.5; DB 2; Length 428;
Best Local Similarity 22.6%; Pred. No. 4.7e-06;
Matches 109; Conservative 59; Mismatches 179; Indels 135; Gaps 21;
QY 18 LDPTTQA-VLGGQFTPMKAAATLMSMLRVDLDRGTVRVLDPGAGVGSITAAALVDRLHTR 76
Db 7 LPPTASGRSLGRVETP--PGLVRFVGLAEARKGVRLFPACADGPF--RAPREA 58
QY 77 PDVAVHVAVETDPFV---PYLRATLEECRNAYGISDIYVEGDYLLNQAGKLDGPFDLV 133
Db 59 HGTGYRFGVGEIDPHALDLPFWAEG-----VWADFLLWEPGE---AFDIL 100
QY 134 IANPPYKGLASDSLARL-----ATTARAVDVPNVVAFVVRVAVISLKEQGRGVF 182
Db 101 LGNPPGIVGEASKYPIHVLRVKGLYKKTSLTWKGYKYNLYGAFIEKSVLLREGGTLVF 160
QY 183 IVPERSWANGPYRQFRHMLMTAVSLDILH---VFESRTKVFA-----DTK 224
Db 161 VVPATWLVDLDFSLRSFLAREGRTEVYILGEVFPGR-KVSAVVLFRKGGKGLALWDTR 219
QY 225 VKQENVI-VAFSVRPOSSSVLSRSVAHGEESIASVPFSALVHDEDDDKIVHFARSASV 283
Db 220 RDGETFTPLLWSEKPEWKGEIIRFETGWTREMEASGPPLGSLFH-----I 264
QY 284 PSAAREFTLADLIGVSTGKVVDLDRGTVRVLDPGAGVGSITAAALVDRLHTR----- 337
Db 265 PFAARSP-----EFKHPAVQKPEP--PGLVPLVTGRNLKPGWIDYESNHSG 309
QY 338 ---PQVGARKPQGFVAVEDVALRQLLPQGSYVVVVKQTAKEDRRRRVIAAVWDGASVALD 394
Db 310 LWWPKERAKELRDFVATPHL-----VVAHTKGTGK-----VVAW 344
QY 395 NKTNYLHESQRPLEKNVAR-----GLMLNLSNVLDQY-----FRAFSGHTQVNGDLRLP 446
Db 345 EKAYPWREEPHLLPKEGVBELDPLFLVWLNLSDKIOEYVKTLYRDFVPHLTRM--LERIP 402
QY 447 FL 448
Db 403 AL 404
RESULT 7
S01615
```

site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) CviBIII - *Chlorella*  
C:Species: *Chlorella* virus CV-NC1A  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S01615  
R:Narva, K.E.; Wendell, D.L.; Skrdla, M.P.; van Etten, J.L.  
Nucleic Acids Res. 15, 9807-9823, 1987  
A:Title: Molecular cloning and characterization of the gene encoding the DNA methyltransferase  
A:Reference number: S01615; MUID:88096540; PMID:3320956  
A:Accession: S01615  
A:Molecule type: DNA  
A:Residues: 1-377 <NAR>  
A:Cross-references: EMBL:X06618; NID:G60638; PIDN:CRA29835.1; PID:G60639  
C:Genetics:  
A:Gene: cvbIIIIM  
C:Superfamily: site-specific DNA-methyltransferase (adenine-specific) CviBIII  
C:Keywords: methyltransferase; restriction modification system; S-adenosylmethionine

Query Match 6.9%; Score 169.5; DB 1; Length 377;  
Best Local Similarity 23.7%; Pred. No. 2.7e-05;  
Matches 75; Conservative 46; Mismatches 123; Indels 73; Gaps 16;

QY 27 GQFFPMKAATLMASMLRVDDLGRGTVRVLDPGAGVGSLSLAALVDRHLHTRPDVAHVAV 86  
DB 26 GIFFPRTVREKLFGEFTEHQNTPGFSLPSCGGEIIECVREF---PLASIKGVEL 81  
QY 87 EDPFVVPVLRATLBECRNAYG-ISYDLVEGDYLLNOGAKLDGPFDLVIANPPY-----G 140  
DB 82 DNDMSTI-----CSKKYAEYNVDIVNEFDLLWKG---GKDFIVGNPPYVRPSSG 128  
QY 141 KLASDLSLARLATTARAVDPVNVVAVFVRAVIS-LKEQGRGVFIVPRSWANGPVRQFRH 199  
DB 129 YKNDNRIRAK-----GRS---NLYVFLYKCIIEHLKEDGILAFIIPSTIGNSFFPEPIRK 180  
QY 200 WLMTAVSLDLHVFBSRTK-VFADTKVQENVIVAFSVRPSQSSVVLRSVAHGEEIAS 258  
DB 181 LIIT---LDLS-FLDLKHDFCDNTR-----LCSIIVKNSPGTKYKT--- 220  
QY 259 SVFPSALVHDEDDDKIVHFAESASVPSAARFLADLIGVSTGKVDFFNRQYLTNDLDA 318  
DB 221 ---YRDIYICDKO-----IPHHGNSYIGSLDLKFKTGFAWANYNK-FFTDKSE- 263  
QY 319 SGVWPMVYQSNIRSGKI 335  
DB 264 ---IPFTSSNKLNEI 277

RESULT 8  
G64577  
type I restriction enzyme M protein - *Helicobacter pylori* (strain 26695)  
C:Species: *Helicobacter pylori*  
C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 17-Mar-2000  
C:Accession: G64577  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
A:Reference number: A64520; MUID:97394467; PMID:9252185  
A:Accession: G64577  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-487 <TOM>  
A:Cross-references: GB:AE000561; NID:G2313564; PIDN:AAD07525.1; PID:G231356  
C:Superfamily: type I site-specific deoxyribonuclease chain hsdM

Query Match 6.7%; Score 165; DB 2; Length 487;  
Best Local Similarity 24.4%; Pred. No. 8.9e-05;  
Matches 63; Conservative 49; Mismatches 90; Indels 56; Gaps 12;

QY 28 QFFTPMKAATLMASMLRVDDLGRGTVRVLDPGAGVGSLSLAALVDRHLHTRPDVAHVAV 87  
DB 208 EYTPLSIASIAKLL-INEFTRNVKIVDSAGTGLMALAHQIGTD---SCTLYAQD 262

QY 88 TDPFVVPVLRATL-----EECRNAYGISYDLVEGDYLLNOGAKLD--GPFDLVIANPPY 139  
DB 263 ISQSLRMLKUNLILNDLTHSLRNA-----IEGNTLTNPVHSKDFGKMDYIVSNPPF 315  
QY 140 GKLASDSLARLATTAR--AVDPN-----VYVAFWVRAVISLKEQGRGVFIVPR 187  
DB 316 KLDPSNEHAEISQNKNDFFLGVPNIPKNSKMPYITLFFQHCLNMLSNKGKGAIIPTG 375  
QY 188 WANGPYRQFRHMLMTAVSLDL-LHVFESR-----TKVPADTKVQENV-IVAFSV 236  
DB 376 FISA-----KSGVENKIIRHLVDRLELVGVVCMPSQVFANTGT---NVSIFQK 422  
QY 237 RPOSSSVLRSVAHGE 254  
DB 423 TFSAKEVVLIDASKLGE 440

RESULT 9  
JU0470  
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) type II - *Acinetobacter*  
N:Alternate names: type II restriction enzyme, M chain  
C:Species: *Acinetobacter calcoaceticus*  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 05-May-2000  
C:Accession: JU0470  
R:Kawakami, B.; Christophe, H.; Nagatomo, M.; Oka, M. Agric. Biol. Chem. 55, 1553-1559, 1991  
A:Title: Cloning and nucleotide sequences of the AccI restriction-modification genes in *Acinetobacter calcoaceticus* strain ATCC 29227  
A:Reference number: JU0469; MUID:91345839; PMID:1368703  
A:Accession: JU0470  
A:Molecule type: DNA  
A:Residues: 1-540 <KAW>  
C:Genetics:  
A:Gene: M-ACCI  
C:Superfamily: site-specific DNA-methyltransferase (adenine-specific) HincII  
C:Keywords: methyltransferase; restriction modification system; S-adenosylmethionine F:123-146/Region: characteristic of N6-adenine methylase

Query Match 6.6%; Score 163; DB 1; Length 540;  
Best Local Similarity 18.7%; Pred. No. 0.00015;  
Matches 95; Conservative 84; Mismatches 176; Indels 154; Gaps 19;

QY 28 QFFTPMKAATLMASMLRVDDLGRGTVRVLDPGAGVGSLSLAALVDRHLHTRPDVAHVAV 87  
DB 31 QFFTPPIAYAMAKWILGNKQLKT--VLEPAFLGVFSRAILSQ-----QKEINIKGFE 82  
QY 88 TDPFVVPVLRATLBECRNAYGISYDLVEGDYLLNOGAKLDGPFDLVIANPPYKGL--- 142  
DB 83 VDETIFENAKYFDDPENY-----NILQDYMTNDWKNKYDG-----IICNPYFKFHDYD 133  
QY 143 ASDSLARLATTARAV--DVPNVVAFWRAVISLKEQGRGVFIVPRSWANGPYRQFRHW 200  
DB 134 NNKILKEITNLKCKLNGTNYTLFLKSIHQSLQNGRCAYIIPSEFLNSDYGLVKTY 193  
QY 201 LMTAVSLDLHVPESKTKVFADTKVQENVIVAFSVRPSQSSVVLRSVAHGEEIAS 260  
DB 194 LIKSKTLRHLIIVDPEENVFDALT-----TASIIIC-----ANDNITDKV 234  
QY 261 PFSALVH-----DEDDDKIVHFAESASV-----PSAARF----- 289  
DB 235 QFNIIQSLQSLDKIDEIINKYPNLFTEQTYNFSNPEIKWKAYQKNSIKFKMLVVF 294  
QY 290 -TLADLIGVSTGKVDFFNRQYLTNDL-----DAGVPMV----- 325  
DB 295 SITAKVVRGIATG-----SNEFTNLKAKFNIDEQYLLPCISAKDAKTSFTTKQD 348  
QY 326 -----YQSNIRSGKID-----WPOVGARKP----- 345  
DB 349 FEELKKSDKSVFLFNAQNSDKNIISSYIQGSEIEINKPFLTASRTPTWYSLNKRKPAP 408  
QY 346 -----QGFAVEDVALRQLLPQGSYVVKRQAKEDRRVIAA--VWDGASVALDNKT 397  
DB 409 VSVNRSGLRFRNEANISNLTSHYCIQNKQVSEIDIDLLFAYLLDTAKQIFEDNSR 468



```

QY      398 NYLHESQRPLEKNVARGIMLWLNSTVLDDQ 426
      : : : : : : : : : : : : : : : : : :
Db      469 QYNGGLQKPEPNDLNKGMWLLDLG--LLDK 495

RESULT 10
XPFS7A
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) Paer7I - Pseudomonas
N:Alternate names: modification methylase Paer7I
C:Species: Pseudomonas aeruginosa
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 05-May-2000
C:Accession: S07366
R:Therriault, G.; Roy, P.H.; Howard, K.A.; Benner, J.S.; Brooks, J.E.; Waters, A.F.; Ging
Nucleic Acids Res. 13, 8441-8461, 1985
A:Title: Nucleotide sequence of the Paer7 restriction/modification system and partial ch
A:Reference number: S07366; MUID:96093653, PMID:3001639
A:Accession: S07366
A:Molecule type: DNA
A:Residues: 1-531 <THE>
A:Cross-references: EMBL:X03274; NID:945385; PIDN:CAA27025.1; PID:gs581435
C:Genetics:
A:Gene: paer7IM
A:Genome: plasmid
A:Start codon: GTG
C:Superfamily: site-specific methyltransferase (adenine-specific) Paer7I
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match      6.6%; Score 162.5; DB 1; Length 531;
Best Local Similarity 23.2%; Pred. No. 0.00016;
Matches 92; Conservative 58; Mismatches 142; Indels 105; Gaps 20;

QY      53 RVLDPCAGVGSUTAAVLVDRL-----HTRPDV--AVHVAVAVETDPFVVPYLRATL 100
      :|:|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      21 RLLEPSFGGGDFLLPIIQRLLSAWRAARPNGTEVDDLGLDAIRAVELHHDTRFTY--AAV 78
      :|:|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |

QY      101 EBCRNAYGSLYD-----LVEGDVLNQGAKLGGPFDLVIANPPY-----GKLAS 144
      :|:|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      79 VALLKREGLSANAATALADWLSSQGFLL---APLEGQDFVVGPNPPYVRPELIPAPLLA 135

QY      145 DSLARLATTARAVDPNVVAVFVWRAVISLKEQGRGVFIVPRSWANGPYVYRQFRHLMTA 204
      :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      136 EYRSRYQT---MYDRADIYPIFIERSILTALSAGNLGFCADRWNKRYGGFLR-----S 187

QY      205 VSLDLHLHFESTKVF---ADTKVKQENVIVAFSVRQSSSVLSRS-----VAHGSE- 254
      :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      188 LVAERPHL-----KVYDVMVDTPAFSDVYATYPAI-----TIISREGGGATIAHRPSI 236

QY      255 -----SIASSVPFSALVHD-----EDDDKIVHFAESASVPSAAREFTL----- 291
      :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      237 DRATLTTLAGLLSAPTLPKDQGPVRELARTVNGAERGCWSLLTKRWLFPAVWRARSHCSKR 296

QY      292 --ADIGIGVSTGKVDFRNRQVLYTD---NLDAAGVVPVMYVQNSIRSGKIDWPQVGARKP 345
      :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      297 LGAREGIGVATG-----ADKAFIGDFESLDVPEPDKLPLVTTKDIMGTEVQMGQGVINP 351

QY      346 QGFVAVEDVALRQLLPQGSVVVVVKRQTAKEDRRRVIA 382
      :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | | :|:| | | | |
Db      352 ----FAESGGLVDL---GEYPLRLRYL--EARDVIA 379

RESULT 11
H71935
type I restriction enzyme (modification chain) - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
C:Accession: H71935
R:Alm, R.A.; Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Werberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923692

```

```
QY 148 --ARLATTARAVDPNVVAVFWRVAVISLKEQGRGVFVPRSWANGPYRFRHLMT-- 203
Db 159 KYRRYTF--IYRADLVVPFIQHSLELLSQGLIGICSDRFTKRYGKKLRKEITDNY 216
QY 204 ---AVSLDILHLHVESRTKVF-ADTKVKQENNVAVFSVRPOSSSVLRSVAHGEESI-- 256
Dd 217 KVRIVDLHKTSPFENEVTAYPAIVIKTKNVKDSV-VRAVTEVITSKVQDADKDFLLS 275
QY 257 ---ASSVPSALVHDE---DDDKIVHFASASVPSAA--RFTLAD-----LGIGV 298
Dd 276 NQKPDQSSKEMKTYVFSWFAGDEPWIIQSQCEIRLRLENRFLIEDDVHSCIRIGV 335
QY 299 STGKVVDPFRNQYLTD---NLDASGVVPMVYQSNIRSGKIDWPQVGARKP-----QGFA 350
Dd 336 ATG-----ADKVIYVDPOVDLEPVLPLVTTADISSGRIIWSKXVINPENSOGGLIN 390
QY 351 VEDV-ALRQLLPQGSVVVVKQTAKED-----RRRVIAAVWDCASRVA 392
Dd 391 LDDFPLKTYFQHEEIIKRNVAKNPQSFQRTIDRIYPEIVHQPKLLIPDMKNTNHTV 450
QY 393 LDNTKNYHESORPL-----EKNVARGMLMLNSTVLDQYFRAFSGHTQVNAAGDLR---- 443
Dd 451 KODGAFYPHNLYIYLPGNWNIDILRAILL---SSVK--FFIWSYATKMRGDTLRYOAO 505
QY 444 -----RLPFLCHREDLILAKVVPDGLPD-----OETLDVAVARLF 478
Dd 506 YLRKIRLP-----DPKSLTNDQKERLMDERVIOQSEYLDLSIVAEIY 546

RESULT 13
I40371
methyltransferase - Bacillus stearothermophilus (fragment)
C:Species: Bacillus stearothermophilus
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 26-May-2000
C:Accession: I40371
R:Rina, M.; Markaki, M.; Bouriotis, V.
Gene 150, 71-73, 1994
A:Title: Sequence of the cloned bseCIM gene: M.BseCI reveals high homology to M.BanIII.
A:Reference number: I40371; MUID:95047552; PMID:7959066
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-579 <RES>
A:Cross-references: EMBL:X79509; NID:9619638; PIDN:CAA56041.1; PID:9619639
C:Genetics:
C:Superfamily: site-specific methyltransferase (adenine-specific) Paer71

Query Match 6.0%; Score 148; DB 2; Length 579;
Best Local Similarity 20.2%; Pred. No. 0.0023;
Matches 92; Conservative 74; Mismatches 162; Indels 128; Gaps 21;

QY 5 ASTETRRQALGKLDPTQAVLGQFFTPMKAATLMSMLRVDDLRG-----TVRVLPDGAG 60
Dd 7 ANTVSQKAT-----GAHFTPDKLAIEVAKRI-LDYFKGKKNRVINLDPACG 53

QY 61 VGSILTAALVDRHLTERPDVAVHVAVETDPFVVPYLRATLEECRNAYGYSYDLVEGDYL- 119
Dd 54 DGEILLAI--NKVAQSMNTQLEIGVDFD---IDAINIANERLSRSGHKNFRLINKDFLE 108

QY 120 -LNOQAKLD-----GPFDLVIANPPYGK---LASDSLARLATTARAVDVPNVVAVFW 168
Dd 109 MWSEGNNDYLDNFTELEFPVDIIIANPPYVTRQILGAEKAKLKEKNLKGKRVLDYQAFV 168

QY 169 RAVISLKEQG-RGVFIVPR--SWANGPYRQFRHLMTAVSLDILHVFESRTKVPADTKV 225
Dd 169 AMTQQLKSGNIIIGVITSNRYLTTKGGGSTRKF-----LVSNFNILIMDLDGSKFFFAAV 223

QY 226 -----KQENVIVAFSVRPOSSSVLRSVAHGEESTASSVPSALVHDE 269
Dd 224 LPAIFFGKKKQKQKNSVNPVKFKIYEQSDI-----EASSVNSFNSLI--- 270

QY 270 DDDKIVHFAESASVPSAAREFTLADLGIGVSTGKVDVFRNRQ-----YLTNLI 316
```

```
Db 271 -----ELLEVNKSLGYSVEDKTYISLIGKIIISPENYKEPWLATDESYEFMKVNO 321
QY 317 DASGVPMVYQSNIRSG-----KIDWPQVGARKPQGFVAVEDVALRQLLPQGSYV 366
Dd 322 NAYGFIEDF--AHVKVGIKTTADSVFRISDNGEL-----PEEQIP-EDKLLRPIISADQ-- 372
QY 367 VVKRQTAKEDRRRVIAAVWDCASRVALDNKNTNYLHE 402
Dd 373 -----ANKW-SVSLVGNKKVLYTHE 392

RESULT 14
F64633
site-specific DNA-methyltransferase (EC 2.1.1.1.-) HP0910 - Helicobacter pylori (strain 26)
C:Species: Helicobacter pylori
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Aug-2000
C:Accession: F64633
R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen-
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: F64633
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-379 <TOM>
A:Cross-references: GB:AE000600; GB:AE000511; NID:92314042; PIDN:AAD07955.1; PID:9231404
C:Genetics:
C:Start codon: TTG
C:Superfamily: site-specific DNA-methyltransferase HP0910
C:Keywords: methyltransferase; restriction modification system

Query Match 6.0%; Score 147; DB 1; Length 379;
Best Local Similarity 28.0%; Pred. No. 0.0015;
Matches 60; Conservative 28; Mismatches 76; Indels 50; Gaps 11;

QY 15 LGKLDPTTQAVLGQFFTPMKAATLMSMLRVDDLRGTVRVLPDGAGVGSILTAALVDRLHT 74
Dd 1 MNNDIKT---LGQVFTPKK---IVDFMLTLKHNGS--VLFPSAGDSGSLKEL----- 46

QY 75 ERPDVAVHVAVETDPFVVPYLRATLEECRNAYGYSYDLVEGDYLLNQAKLDGPPDLVI 134
Dd 47 -----KKAVRIEIDPKICP-----KNALCMDF----FDY-----PLENQPDITII 81

QY 135 ANPPYGK---LASDSLARLATTARAVDVPNVVAVFWRVAVISLKEQGRGVFVPRSWANG 191
Dd 82 GNPPYVVKHDIAPSTKEKLHYSL-FDERSNLYLFFIEKAIKHLKPKGELIFITPROFLKS 140

QY 192 PYYRQFRHLMTAVSLDILHVFESRTKVPADTKV 225
Dd 141 TSSVKLNWEIYKEGT--ITHFFE-----LGQDKV 167

RESULT 15
H97333
site-specific modification DNA-methyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 01-Mar-2002
C:Accession: H97333
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C.
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97333
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK81459.1; PID:gl5026627; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
```



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 11.4141 Seconds  
(without alignments) 2217.092 Million cell updates

Title: US-10-668-047-2

Perfect score: 2469  
Sequence: 1 VHPIASTETRRQAAALGKLDP.....QETLDAVVARLFECEIPSS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query %				DB	ID	Description	
	Score	Match	Length					
1	667.5	27.0	507	1	MTPS	PROST	P00474	providencia
2	662.5	26.8	501	1	MTBA	THEAU	P33563	bacillus su
3	183	7.4	421	1	MTB8	THESQ	P14385	thermus aqu
4	172.5	7.0	428	1	MTB8	THETH	P29749	thermus the
5	171.5	6.9	587	1	MTS1	SPRAL	Q53609	streptomyce
6	169.5	6.9	377	1	MTG3	CHVNI	P10835	chiorella v
7	167	6.8	540	1	MTAL	ACICA	P25201	acinetobact
8	162.5	6.6	549	1	MTP7	PSEAE	P55103	pseudomonas
9	151	6.1	561	1	MTV1	BACST	P43422	bacillus st
10	148	6.0	579	1	MTV1	BACST	P43423	bacillus st
11	138.5	5.6	527	1	MTV1	YANCR	P96188	xanthomonas
12	134	5.4	580	1	MTB3	BACAR	P22772	bacillus an
13	132	5.3	520	1	TIMP	ECOLI	Q47163	escherichia
14	132	5.3	1181	1	YYO2	METJA	Q60301	methanococc
15	130	5.3	294	1	KSGA	OCEIH	P59155	o dimethyla
16	130	5.3	520	1	TIM1	SCOLI	P10484	escherichia
17	128	5.2	502	1	MTHC	HAEN	P17444	haemophilus
18	128	5.2	518	1	MTHD	HAEN	P44414	haemophilus
19	125.5	5.1	317	1	KSGA	MYCTU	O05595	m dimethyla
20	118	4.8	290	1	KSGA	STRPN	O97nn5	s dimethyla
21	118	4.8	291	1	KSGA	STRMU	P59156	streptococc
22	118	4.8	1372	1	RPOC	RICCN	Q9rh40	rickettsia
23	117	4.7	294	1	KSGA	LACLA	Q9chn8	lactococcus
24	116.5	4.7	245	1	YFIC	ECOLI	P31825	escherichia
25	115.5	4.7	271	1	KSGA	METWA	O8pu18	methanosarc
26	115.5	4.7	273	1	KSGA	METHH	O27381	methanobact
27	115.5	4.7	544	1	MTS7	ECOLI	P25240	escherichia
28	115.5	4.7	558	1	TIMH	METJA	Q60297	methanococc
29	115	4.7	290	1	KSGA	STRA3	O8e3d7	s dimethyla
30	115	4.7	292	1	HEMK	HAEN	P45253	haemophilus
31	114	4.6	290	1	KSGA	STRR6	Q8dnd3	s dimethyla
32	113	4.6	295	1	PRNA	BRWA	O89fw1	bradyrhizob
33	112	4.5	290	1	KSGA	STRA5	Q8dxx8	s dimethyla

## ALIGNMENTS

## RESULT 1

```

MTPS_PROST
ID   MTPS_PROST          STANDARD;          PRT;   507 AA.
AC   P00474;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Modification methylase PstI (EC 2.1.1.72) (Adenine-specific
DE   methyltransferase PstI) (M.PstI).
DE   PSTM.
GN   Providencia
OS   Providencia stuartii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Providencia.
NCBI_TaxID=588;
[1]
RN   SEQUENCE FROM N.A.
RP   STRAIN=164;
RC   MEDLINE=84239756; PubMed=6330092;
RX   Walder R.Y., Walder J.A., Donelson J.E.;
RT   "The organization and complete nucleotide sequence of the PstI
RT   restriction-modification system.";
RL   J. Biol. Chem. 259:8015-8026(1984).
CC   -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC   CTGAG. CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC   PROTECTS THE DNA FROM CLEAVAGE BY THE PSTI ENDONUCLEASE.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC   adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   -!- SUBUNIT: Monomer.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/
CC   or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; K02081; AAA25672.1; -.
DR PIR; A00553; XYOFS.
DR DR REBASE; 3483; M.pstI.
DR DR InterPro; IPR002296; N12N6_mtfrase.
DR DR InterPro; IPR002052; N6_Mtase.
DR DR InterPro; IPR000051; SAM_bind.
DR DR PRINTS; PRO0507; N12N6MTFRASE.
DR DR PROSITE; PS00092; N6_MTASE; 1.
DR DR TRANSFASER; MethyItransferase; Restriction system.
DR DR SEQUENCE 507 AA; 56877 MW; AF5DA9341741B23D CRC64;
SQ

```

Query Match	27.0%;	Score 667.5;	DB 1;	Length 507;
Best Local Similarity	34.2%;	Pred. NO. 2.7e-41;		
Matches 166;	Conservative	87;	Mismatches 206;	
			Indels	27.
				Cons.

QY	11	ROAALGKLDPTTQAVLGQFFTPMKAATLASMLRVDDLRGTIVRVDLPGAGVGSILTAALVD	70
		: :         : :         : :	
Db	20	REAA NSTLD ETRSKLGQPMS SSVSELMA NLF - ESYVGEHEI LDAGAGVGSILTA AFVQ	77

```

QY 71 RLHTEPDVAVHVVAVETDPFVVPVYLRATLEECR---NAYGISYD--LVEGDYLLNQCAK 125
DB 78 NA-TLNGAKSISSTCYEISEVMVYNLIQVLDLCKIRAMEFEVNMWOCKIIESDFIQASVEQ 136
QY 126 L-----DGPFDLVIANPPYKGLASDLARLATTARAVDVPVNVYAFWVRVAVISKEQGRG 180
DB 137 LLIENYSKYKAILNPPYKIAKGRREALLQKVGIEASNLISAFVALAIKQLSGGEL 196
QY 181 VFIIVRSWANGPYRQFRHMLMTAVSLDILHVFESEKTKVFADTKVKQENVIIVAFSVRQFS 240
DB 197 VAITPRSCFNGFYFNDKQMLDECNLKIHVFNRSKSAFAKADNVLQENIIYHITKGETQ 256
QY 241 SSV--VLSRVAHGSESTASSVPFSALVHDEDDKIVHFA-----ESASVPSAARFTLA 292
DB 257 RKWTVYYSTCANDINPTIFEVFPDEIVKSNPNPDLFHIVTNEQRELANKAGGLPCSL 316
QY 293 DLGIGVSTGKVVDFFNRVLTDNLDSAGVPMVYQSNIRSGKIDPQVGARKPQGFVAVE 352
DB 317 DLGIVSTGKVVDFFNRVLTDNLDSAGVPMVYQSNIRSGKIDPQVGARKPQGFVAVE 374
QY 353 DVALRQLLPQGSYVVVVKQTKAKEDRRRVIAAVW----DGASVALDNKTNVHESQRPLE 408
DB 375 EATNLMVNGIYVLTRELTKAEKRRIVASTIYYPDIANVDTVGFDNKINYFHANGKPLD 434
QY 409 KNAVGLMLNLTSLVDQVAFSGHTQVAGDLRLPFLCREDLILLAKVVPDGLPQOE 468
DB 435 ISLAKGLMWFLNSTLIDKYFROMNGHTQVATDLRALYPTREQLIEDIANQVDFGEFOT 494
QY 469 TLDVAV 474
DB 495 KIDEII 500

RESULT 2
MTTB BACSU
ID MTTB BACSU STANDARD; PRT; 501 AA.
AC P14385;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Modification methylase BsuBI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase BsuBI) (M.BsuBI).
GN HSDBM OR HSBM
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISB8;
RX MEDLINE=93126092; PubMed=1480472;
RA Xu G.-L., Kapfer W., Walter J., Trautner T.A.;
RT "BsuBI -- an isospecific restriction and modification system of PstI:
RT characterization of the BsuBI genes and enzymes.";
RL Nucleic Acids Res. 20:6517-6523(1992).
CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CTGACG, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRAND, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE BsuBI ENDONUCLEASE.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L01541; AAA18169.1; -.
CC DR PIR; S35515; S35515.
CC DR REBASE; 3336; M.BsuBI.
CC DR InterPro; IPR002296; N12N6 mtfrase.
CC DR InterPro; IPR002052; N6_Mtase.

```

```

DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Methyltransferase; Restriction system.
SQ SEQUENCE 501 AA; 57197 MW; AC33B2944699653E CRC64;

Query Match 26.8%; Score 662.5; DB 1; Length 501;
Best Local Similarity 36.4%; Pred. No. 6.1e-41;
Matches 173; Conservative 80; Mismatches 193; Indels 29; Gaps 12;

QY 26 LQGFPTPKMAATLMASMLRVDDLRCTVVLDPAGAGVSGISITAAALVDRLHTEPDVA-VHV 84
DB 25 LQGFPTPSISIFWACLFSDEKL--NNAKVLDAAGAGISLTSATLRLISENIGKADLHI-- 82
QY 85 AVETDPFVVPVYLRATLEECRNAYGISYDLVEGDY-----LLNQG---AKLDGPFDLVI 134
DB 83 -LEIDEMLEPYLSETLALFKDYIEINSQIIDDFFIEMAAYSLLDEESLAKDKQRFTHAI 141
QY 135 ANPPYKGLASDLARLATTARAVDVPVNVYAFWVRVAVISKEQGRGVFIVPRSWANGPY 194
DB 142 LNPPYKTKSNSKHKRLRKAGIETVNLISAFVALTVDLMSDGGEIVFIIPSCFNGPYV 201
QY 195 RQFRHMLMTAVSLDILHVFESEKTKVFADTKVKQENVIIVAFSVRQSSSVLSRSVAHGE 254
DB 202 RFRHMLMTAVSLDILHVFESEKTKVFADTKVKQENVIIVAFSVRQSSSVLSRSVAHGE 261
QY 255 SIAS-SVPPSALVHDEDDKIVHFAESASVPSAAR-----FTLADLIGVSTGKVVDFFNR 308
DB 262 VTRSVRYPEKIVQPNDELEKFIHINTNEETLIEKHPNVCSLEELNIEVSTGVPVDFRV 321
QY 309 RQYLTNDLDSAGVPMVYQSNIRSGKIDPQVGARKPQGFVAVEVALRQLLPQGSYVVV 368
DB 322 KENLRE-MPGECTVPLFYPNHFVGTSLSEYPRKM-MKKPNAILIRNEKVE-KWLYPNGHYVVV 378
QY 369 KQQTAKEDRRRVIAAVWDGAS---RVALDNKTNVHESQRPLEKNVAGLMLNLTSLV 424
DB 379 KFPSKKEKRIRVAGVLTPEVNDDPVGVFENGVLNHLNYSKISGEVAGLYALNLTPTV 438
QY 425 DOYFRAFSGHTQVAGDLRLPFLCREDLILLAKVVPDGLPD--QETLDVAVVARL 477
DB 439 DKYFRIFNGHTQVATDLRLTKMFPSPDILISLKGWNIENIVNGVQVEIDSKLEEL 493

RESULT 3
MTTB THEAQ
ID MTTB THEAQ STANDARD; PRT; 421 AA.
AC P14385;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Modification methylase TaqI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase TaqI) (M.TaqI).
GN TAQIM.
OS Thermus aquaticus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OX NCBI_TaxID=271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YTI1;
RX MEDLINE=88096538; PubMed=2827113;
RA Slatko B.E., Benner J.S., Moran L.S., Jager-Quinton T., Simcox T.G.,
RA van Cott E.M., Wilson G.G.;
RT "Cloning, sequencing and expression of the Taq I restriction-
RT modification system.";
RL Nucleic Acids Res. 15:9781-9796(1987).
RN [2]
RP REVISIONS.
RC STRAIN=YTI1;
RX MEDLINE=92201707; PubMed=1551602;
RA Barany F., Slatko B., Danzitz M., Cowburn D., Schildkraut I.,
RA Wilson G.G.;
RT "The corrected nucleotide sequences of the TaqI restriction and
RT modification enzymes reveal a thirteen-codon overlap.";

```

RL Gene 112:91-95(1992).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).

RX MEDLINE=95062184; PubMed=7971991;

RA Labahn J., Granzin J., Schluckebier G., Robinson D.P., Jack W.E.,

RA Schildkraut I., Saenger W.;

RT "Three-dimensional structure of the adenine-specific DNA

RT methyltransferase M.Taq I in complex with the cofactor S-

RT adenosylmethionine.";

RT Proc. Natl. Acad. Sci. U.S.A. 91:10957-10961(1994).

CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE

CC TCGA, CAUSES SPECIFIC METHYLATION ON A-4 ON BOTH STRANDS AND

CC PROTECTS THE DNA FROM CLEAVAGE BY THE TAAI ENDONUCLEASE.

CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-

CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M76681; AAA27506.1; -.

DR EMBL; Y00499; CAA68551.1; ALT\_SEQ.

DR PIR; JN0257; JN0257.

DR PDB; 2ADM; 27-JAN-97.

DR PDB; 1AQL; 12-FEB-97.

DR PDB; 1AQJ; 12-FEB-97.

DR PDB; 1G38; 12-MAR-01.

DR REBASE; 3520; M.TaqI.

DR InterPro; IPR002296; N12N6 mtfrase.

DR InterPro; IPR002052; N6 Mtase.

DR InterPro; IPR000051; SAM bind.

DR PRINTS; PR00507; N12N6MTFRASE.

DR PROSITE; PS00092; N6\_MTASE; 1.

KW Transferase; Methyltransferase; Restriction system; 3D-structure.

FT HELIX 25 34

FT TURN 39 40

FT STRAND 42 46

FT TURN 47 47

FT TURN 49 50

FT HELIX 52 61

FT STRAND 66 71

FT TURN 74 76

FT TURN 81 82

FT STRAND 84 87

FT TURN 90 92

FT STRAND 99 104

FT HELIX 122 132

FT TURN 134 135

FT TURN 138 139

FT HELIX 142 152

FT TURN 153 153

FT STRAND 154 165

FT HELIX 166 169

FT TURN 170 170

FT HELIX 172 174

FT HELIX 175 184

FT STRAND 186 193

FT TURN 199 200

FT STRAND 203 209

FT TURN 215 223

FT STRAND 224 225

FT TURN 226 235

FT TURN 236 237

FT HELIX 248 255

FT TURN 256 256

FT STRAND 257 259

FT HELIX 260 263

FT STRAND 264 264

FT HELIX 272 276

FT TURN 277 277

FT TURN 279 280

FT STRAND 282 282

FT TURN 287 288

FT STRAND 289 291

FT STRAND 293 293

FT HELIX 295 297

FT STRAND 298 299

FT TURN 300 301

FT STRAND 302 303

FT STRAND 313 315

FT HELIX 316 321

FT TURN 322 322

FT HELIX 324 327

FT STRAND 330 333

FT STRAND 343 346

FT TURN 347 348

FT STRAND 352 352

FT STRAND 356 360

FT TURN 362 363

FT STRAND 364 366

FT HELIX 368 375

FT TURN 376 376

FT HELIX 378 388

FT TURN 389 390

FT HELIX 397 400

FT TURN 401 402

FT STRAND 404 406

FT TURN 407 409

FT STRAND 410 412

SQ SEQUENCE 421 AA; 47848 MW; 6F7D2081EB7FCE45 CRC64;

Query Match 7.4%; Score 183; DB 1; Length 421;

Best Local Similarity 23.7%; Pred. No. 5.1e-06;

Matches 108; Conservative 58; Mismatches 187; Indels 102; Gaps 20;

Qy 26 LGQFTPMKAATLWASMLRVDDLGRGTVRLDPCGAGVCSLTAALVDRLHTEPDPVAVHVA 85

Db 18 LGRVETPEVVDVFWSLAEAP--RGG-RVLEPCAHGPFLL-----RAFREAHGTGYRFVG 69

Qy 86 VETDPFVW---PYIRATPLEECRNAYGISYDLVEGYLLNQAKLDGPFDLVIANPPYG-- 140

Db 70 VEIDPKALDPPWAEGIL-----ADFLWEPGE---AFDLILGNPPYGIV 111

Qy 141 -----KLASLSLARLATARADVDPNVVAFWVRAVISLKEQGRGVFTVPRSWA 189

Db 112 GEASKYPIHVFKAVKOLYKKAFTWKG--KYNLYGAFLEKAVRLKPGGVLFVVPATWL 169

Qy 190 NGPYRQFRHWMATVSLDILHVFESRTKYFADTKYQENVIVAFSVRPOSSSVLSRSV 249

Db 170 VLEDFALLREFLAR-----EGKTSVYIYGEVFPQKKVSAVVIRFOKSGKLSLWD 219

Qy 250 AHGEESIASVPFSAVLVHDDDDKIYHF-AESASVPSAARFTLADLIGVSTGKVVDFRN 308

Db 220 TQESSESGFTPLWAEYPHWEGE--IIRFETEETKLEISGMPGLDLFHPFAARSPEFKK 277

Qy 309 RQYLTNLDASGVVPMVYOSNIRSGKIDW-----POVGARKPGFVAVEDVALRQL 359

Db 278 HPAVRKE-PQGLVPLVTGRNLKPGWVDYKNSGLWMPKERAKELRDFVAT----- 328

Qy 360 LPOGSSVYVVKROTAKEDRRRIVAAVWDGASRVALDNKTNVYLHESQRPLEKQVAR----GL 415

Db 329 -----PHLVV-----AHTKGRVVA--W-----DERAYFWREBEFHLKPEGVRLDPSSL 371

Qy 416 MLWLNSTVLDQY----FRAFSGHTQVNAAGDLRLRP 446

Db 372 VQWLNSEANQKHVRTLYRDFVPHLTIRM--LERLP 404

RESULT 4

MTT8 THETH

ID MTT8 THETH

AC P29749;

STANDARD; PRT; 428 AA.





Db 138 HHLGIDQKRD--VARCADIQIKPSGLSGLYLVFVLLSHLLRADAVSTWLPSEFMDV 195  
Qy 192 PYRQFRHLMTAVSLDILHVFESTKVFADTKVKQENVIWAFSVRPOSSSVLSRS--- 248  
Db 196 NYGTALKEYLATRVOLVRIHQVDAEVOFDDALV-----TSSVVVFRNSPP 241  
Qy 249 -VAH-GEBSIASVFPFSAHVDEDDDKIVHFAESASVPSAARFTL-----ADIGIGVS 299  
Db 242 RPKHTABFSFGTL-----SEPKVTHQIPSAALTPKAKWSRYTVGVMPADINLKQT 292  
Qy 300 TGKVVD--PRNROYLTNDLDSGVVPMVYQSN-----RSKID----- 336  
Db 293 GPKLSDFFKIRGLATSNFAFIIPRSEARLIGIKRNLFRPLPSPRKLKGDATADASG 352  
Qy 337 WPOVGAR-----KPOGFVAVEDVALRQLLP-----QGSYVVVVKQT-AKEDRRR--V 380  
Db 353 WPDPEQALLDCPLPIEDLLENPALAAYLSTADEKIRGGYLVSKSPWTKQOREPAP 412  
Qy 381 IAAVWDGASVALDNKNYLNHESORPLE--KN-----VARGMLMLNSTVLQYFRAFSGH 434  
Db 413 ILTYMGRGK-----DDQHLRFIRNDSDAVATNMYLMLYPTALLQRYL----- 456  
Qy 435 TQVWAGDLRLPFLCREDLILLA-----XVPDGLPDQF-----TLDAVV 474  
Db 457 ----AGDPERIKQVHKALLAITAADLRGGGRVYGGGLHKMEPKELAAPADGIATLDPVL 512

## RESULT 6

MT3 CHVN1  
ID MT3 CHVN1 STANDARD; PRT; 377 AA.  
AC P10835;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Modification methylase CviBIII (EC 2.1.1.72) (Adenine-specific  
methyltransferase CviBIII) (M.CviBIII).  
GN CviBIII.  
OS Chlorella virus NC-1A (CV-NC1A).  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.  
OX NCBI\_TaxID=10504;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88096540; PubMed=3320956;  
RA Narva K.E., Wendell D.L., Skrdla M.P., van Etten J.L.;  
RT "Molecular cloning and characterization of the gene encoding the DNA  
methyltransferase, M.CviBIII, from Chlorella virus NC-1A.",  
RL Nucleic Acids Res. 15:9807-9823(1987)  
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
TCGA, CAUSES SPECIFIC METHYLATION ON A-4 ON BOTH STRANDS.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X06618; CAA29835.1; -  
DR HSP; P14385; 1AQU.  
DR REBASE; 752; M.CviBIII.  
DR InterPro; IPR002296; N12N6\_mtfrase.  
DR InterPro; IPR002052; N6\_Mtase.  
DR PRINTS; PR000051; SAM\_bind.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
XQ Transferase; Methyltransferase.  
SQ SEQUENCE 377 AA; 42895 MW; B22471367D61BBC CRC64;

Query Match

Best Local Similarity 6.9%; Score 169.5; DB 1; Length 377;

Pred. No. 4.3e-05;

Matches 75; Conservative 46; Mismatches 123; Indels 73; Gaps 16;  
Qy 27 GQFTPKMAATLMASMLRVDDLRTGTVRLVDPGAGVGSITAAALVDRLHTRPDVAHVAV 86  
Db 26 GIFFTPKTVREKLFQGFTEHFQNTPGFSILEPSCGTGEIISECVERF-----PLASIKGVEL 81  
Qy 87 ETDPFVVVYLRATLEECRNAG--ISYDLVEGDYLLNOGAKLDGPPDLVIANPPY-----G 140  
Db 82 DNDSTI-----CSKYAEYNVDVINEDFLLWG-----GRFDFIVGVPPYVVRPSG 128  
Qy 141 KLASDSLARLATTARAVDVNPNVYVAVVRAVIS--LKEQGRGVFVPRSWANGPYVQRPH 199  
Db 129 YKDNRIAK-----GRS-----NLYVEFLYKCIHELKEDGILAFIIPSTIGNSSFYEPIRK 180  
Qy 200 WLMTAVSLDILHVFESTK-VFADTKVKQENVIWAFSVRPOSSSVLSRSVAHGEESTAS 258  
Db 181 LIIT--LDLIS-FEILDKHD FCDTNR-----LCSIVIKNSPGTKYT--- 220  
Qy 259 SVFPSALVHDEDDDKIVHFAESASVPSAARFTLADLGIVSTGKVVDFRNRQYLTNDLDA 318  
Db 221 ---YRDYICXD-----IPHGNSYIGSLDKFKTGFAMANVNK-FTFDKSE- 263  
Qy 319 SGVVPVYQSNIRSGKI 335  
Db 264 ---IPFTSSNIKINEI 277

## RESULT 7

MTAL ACICA  
ID MTAL ACICA STANDARD; PRT; 540 AA.  
AC P25201;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Modification methylase AccI (EC 2.1.1.72) (Adenine-specific  
methyltransferase AccI) (M.acCI).  
GN ACCIM.  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Moraxellaceae; Acinetobacter.  
OX NCBI\_TaxID=471;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=ATCC 49823;  
RX MEDLINE=91345839; PubMed=1368703;  
RA Kawakami B., Hilzheber C., Nagatomo M., Oka M.;  
RT "Cloning and nucleotide sequences of the AccI restriction-modification  
genes in Acinetobacter calcoaceticus.",  
RL Agric. Biol. Chem. 55:1553-1559(1991).  
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
GTMKAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND  
PROTECTS THE DNA FROM CLEAVAGE BY THE ACCI ENDONUCLEASE.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
CC -!- SUBUNIT: Monomer.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D10671; BAA01523.1; -  
DR REBASE; 3271; M.ACCI.  
DR InterPro; IPR002296; N12N6\_mtfrase.  
DR InterPro; IPR002052; N6\_Mtase.  
DR PRINTS; PR000051; SAM\_bind.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
KW Transferase; Methyltransferase; Restriction system.  
SQ SEQUENCE 540 AA; 63092 MW; A2569693712E0F59 CRC64;

```
Query Match      6.8%; Score 167; DB 1; Length 540;
Best Local Similarity 18.9%; Pred. No. 0.00011;
Matches 96; Conservative 84; Mismatches 175; Indels 154; Gaps 19;

QY 28 QFFTPMKAATLWASMLRVDDLTGTVRVLDPCGAGVGSILTAALVDLHTRPDPVAVHVAVE 87
DB 31 QFFTPFPPIAYAMAKWILGNKQKLT-VLEPAFGLGVFSRILSQ-----QKEINIKAFE 82
QY 88 TDPFVVPYLRATLECRNAYGISVDLVEGDYLLNQA-KLGGPDLVIANPPYKGL----- 142
DB 83 VDEITIFENAKYFDDFENV-----NILLQDYNDWKNKYDG-----IICNPPYKFDHYD 133
QY 143 ASDSLARLATTARAV--DVPNVYVAFWRAVISLKEQGRGVFIIVPRSWANGPYVQPRHW 200
DB 134 KNNILKETETNLCKKLNGFTNLTLLFLLSIHQSNGRCAYIIPSEFLNSDYGLVKTY 193
QY 201 LMTAVSLDILHVFESRTKVFADTKVKQENVIVAFSVRPSQSSVLSRSVAHGEESIASSV 260
DB 194 LIKSKYLRHIIIVDEENVFDALT-----TASIIIC-----ANDNITDKV 234
QY 261 PFSALVH-----DEDDKKIVHFAESV-----PSARF----- 289
DB 235 QFNNIQSQDLSKIDELINKVPNFLETEQTYNFSDLNPEIKWAKYQKNSIKFKNLVPF 294
QY 290 -TLADLIGVSTGKVVDFNRQYLTDLN-----DASGVVPMV----- 325
DB 295 STYAKVVRGIATG-----SNEYFTFNLSKAKEFNIDEQYLUPCISAKDAKTSFTKQD 348
QY 326 -----YQSNIRSGKID-----WPGVGARKP----- 345
DB 349 FEELKSKSVFLFNAQNSTDKNISYIQGSEINKRFLTASRTPMVSYLENRPAPIW 408
QY 346 -----QGFVAVEDVALRQLPQGSYVVVVKROTAKEDRRVIAA--VWDGSRVALDNKT 397
DB 409 VSVFNRSGLRFTIRNEANISLTSYHCIIQNKQVSVSEIDILFAYLLTDTAKQIFEDNSR 468
QY 398 NYLHESQRLPKLVNARGMLNLSVTLQD 426
DB 469 QYGNGLQKPEPNDLKNKGMMLDLG--LLDK 495

RESULT 8
MTP7_PSEAE STANDARD; PRT; 549 AA.
AC P05103;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Modification methylase Paer7I (EC 2.1.1.72) (Adenine-specific
DE methyltransferase Paer7II) (M.Paer7I).
GN PAER7IM.
OS Pseudomonas aeruginosa.
OG Plasmid pMG7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OC NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86093653; PubMed=3001639;
RA Theriault G., Roy P.H., Howard K.A., Benner J.S., Brooks J.E.,
RA Waters A.F., Gingers T.R.;
RT "Nucleotide sequence of the Paer7 restriction/modification system and
RT partial characterization of its protein products.";
RL Nucleic Acids Res. 13:8441-8461 (1985).
RN [2]
RP REVISIONS TO N-TERMINUS.
RX MEDLINE=95331587; PubMed=7607512;
RA Timinskas A., Butkus V., Janulaitis A.;
RT "Sequence motifs characteristic for DNA [cytosine-N4] and DNA
RT [adenine-N6] methyltransferases. Classification of all DNA
RT methyltransferases.";
RL Gene 157:3-11 (1995).
```

```
-!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CTCAG, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
PROTECTS THE DNA FROM CLEAVAGE BY THE PAER7I ENDONUCLEASE.
-!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
adenosyl-L-homocysteine + DNA 6-methylaminopurine.
-!- SUBUNIT: Monomer.
-!- This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL: X03274; CAA27025.1; ALT_INIT.
HSPF: P14385; 2ADM.
REBASE: 3477; M.Paer7I.
InterPro: IPR002296; N12N6_mtfrase.
DR PRINTS: PR00507; N12N6MTFRASE.
DR PROSITE: PS00092; N6_MTASE.1.
KW Plasmid; Transferase; Methyltransferase; Restriction system.
SQ SEQUENCE 549 AA; 61160 MW; 3A9A5409A25BA092 CRC64,
Query Match      6.6%; Score 162.5; DB 1; Length 549;
Best Local Similarity 23.2%; Pred. No. 0.00023;
Matches 92; Conservative 58; Mismatches 142; Indels 105; Gaps 20;

QY 53 RVLDPCAGVGSILTAALVDRL-----HTEPDEV-AHVVAVEVDPFVVPYLRATL 100
DB 39 RLLEPFGGGDFLLPIIQRLLSAMRAARPNGTEVDLDGAIHRAVELHDDTFRSTY--AAV 96
QY 101 BECRNAYGISYD-----LVEGDYLLNQAKLDGPFDLVIANPPY-----GKLAS 144
DB 97 VALLKRGISANAATALADRWLSQGFLL--APLEGQDFVVGVPYVRPELIPAPLLA 153
QY 145 DSLARLATTARAVDVPNVYVAFWRAVISLKEQGRGVFIIVPRSWANGPYVQPRHWLMTA 204
DB 154 EYRSRYQT--MYDRADYIPIETSLTALSAGNLGFCADRWKMKRYGGFLR-----S 205
QY 205 VSLDILHVFESRTKVP--ADTKVKQENVIVAFSVRPSQSSVLSRS-----VAHGE- 254
DB 206 LVAERFHL-----KVYVDMVDTPAFHSDVIAPAI-----TIISREGGATRIARPSI 254
QY 255 -----SIASSVPFSALVHD---EDDDKIVHFAESASVPSAARFTL----- 291
DB 255 DRATLTLLAGLSAPTLPKDAGVRELARTNGAERGCSLLTRWLFVWRAHSCSKR 314
QY 292 --ADLIGVSTGKVVDFNRQYLT-----NLDASGVVPMVYQSNIRSGKIDPQVGARKP 345
DB 315 LGARFGIGVATG---ADKAFIGDPESLDVBPDRKPLVLTDTKIMTGQVWRGQGVINP 369
QY 346 QGFVAVEDVALRQLPQGSYVVVVKROTAKEDRRVIA 382
DB 370 -----FAESGGLVDL---GEYPLRLRYL--EARRDVIA 397

RESULT 9
MTVL_BACST STANDARD; PRT; 561 AA.
ID MTVL_BACST
AC P43422;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Modification methylase BstVI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase BstVI) (M.BstVI).
GN BstVIM.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
```



```
QY 317 DASGVVPMVYQSNIRSG-----KIDMPQVGARKPQGVAVEDVALRQLLPQGSYV 366
Db 322 NAYGFIEDF--AHVKGVIKTTADSVFIRSDMGEL-----PEQIP-EDKLLRPIISADQ-- 372
QY 367 VVKRQTAKEDRRRVIAVWDGASRVALDNKTKNYLHE 402
Db 373 -----ANKW-SVSLGVNKKVLYTHE 392

RESULT 11
MTX1_XANCR
ID MTX1_XANCR STANDARD; PRT; 527 AA.
AC P96188;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase XamI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase XamI) (M.XamI).
GN XAMIM.
OS Xanthomonas campestris (pv. amaranthicola).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=54735;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 11645;
RX MEDLINE=97276896; PubMed=91305899;
RA Gomez P., Ribas-Aparicio R.M., Pelaez A.I., Gomez A., Rodicio M.R.;
RT "Isolation and nucleotide sequence of the gene encoding the XamI DNA
RT methyltransferase of Xanthomonas campestris pv. amaranthicola.";
RL Biochim. Biophys. Acta 1351:261-266 (1997).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC GTGCAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE XAM I ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U77781; AAD13686.1; -.
CC REBASE; 3526; M.XamI.
CC InterPro; IPR002296; N12N6_mtfrase.
CC InterPro; IPR002052; N6_Mtase.
CC InterPro; IPR000051; SAM_bind.
CC PRINTS; PR00507; N12N6MTFRASE.
CC PROSITE; PS00092; N6_MTASE; FALSE_NEG.
CC TRANSFERASE; Methyltransferase; Restriction system.
CC KW SEQUENCE 527 AA; 57535 MW; B80C43CE684D8956 CRC64;

Query Match 5.6%; Score 138.5; DB 1; Length 527;
Best Local Similarity 21.5%; Pred. No. 0.012;
Matches 111; Conservative 68; Mismatches 204; Indels 133; Gaps 25;

QY 15 LGKLDPTTQAVLGQFFTPMKAATLMASMLRVDLDLRT-VRVLDPGAGVSLTAALVDRHL 73
Db 66 LGHRAFRAGAV---YTP---AFIVSRMTWLAAQSPARIVDPGAGSRFILAAGEAF- 117
QY 74 TERPDVAHVAVETDPFVVYPYLRATLECRNAYGISYDLVEGDYLLNQCAKLD--GPPD 131
Db 118 ---PD---AQIVAVEMDPLAALMLRANLS---ARGMTDRATVMVKDY---REVKLPPCAGIT 167
QY 132 LVIANPYGK---LADSLARLATT-----ABAVDVPNVVAFVRAVISLKEQGRGVFI 183
Db 168 AFTGNPPYVKKHHDIGEDWKAWYASNFAGYGIKASALAGLHLHFLOTRLLAKAGDVGAFT 227
QY 184 VPRSWANGPYRQFRHLMVMTAVSLDILHVFESRTKVFADTKVKQENVIVAFSV-----RPQ 239
```

```
Db 228 TSAEMWDMVYGSALRRLRLDLLDELGGIALHVLPEPTVEAFPGTATTA--AIACFRVGETARPV 285
QY 240 SSSVV-----LSRSVAHGEESIASVPPFSALVHDEDDDKIVHFAESAGSVPSAARFTL 291
Db 286 RVRFIDELTNLNGITKGTDIPEQLQAASRSLIVR-----PSAPAMA----- 328
QY 292 ADLGIG-----YSTGKVDFRNRYQLTDNLDSGVVPMVYQSNIRSGKIDWFOVGARKPQG 347
Db 329 GDIELGELFRVHRGQVTG-ANGIWIAGE-HAQGLPDRVMPAVTKAK-DLIQAAGAHLSA 385
QY 348 FVAVEDVALRQLLPQGSVYVVKRQTAKEDRRVIA-----AVWDGASRVALDNKTKNYLHES 403
Db 386 EV-----LRRVIDLPTDL-----DDFTKEERRRISFLSWAKLHGADQ-----SYIAQH 429
QY 404 QR-----PRT; 580 AA.
Db 430 RRAWWSVGLKAPAPILCTYMARPPQFTLNACDARHINIAHGLYPREPLAAGIMASLVTW 489
QY 419 LNSTVLDOYFRAFSGH-TQVNAGDLRRLPFLCREDL 453
Db 490 LNKNIINTSGRTYAGGLTKFEPKEIERLRIPSLLENL 525

RESULT 12
MTB3_BACAR
ID MTB3_BACAR STANDARD; PRT; 580 AA.
AC P22772; O44656;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase BanII (EC 2.1.1.72) (Adenine-specific
DE methyltransferase BanII) (M.BanII).
GN BANIIIM.
OS Bacillus aneurinolyticus.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae;
OC Aneurinibacillus group; Aneurinibacillus.
OX NCBI_TaxID=1391;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91242093; PubMed=1368640;
RA Kawakami B., Sasaki A., Oka M., Maekawa Y.;
RT "Nucleotide sequence of the gene coding for the BanIII DNA
RT methyltransferase in Bacillus aneurinolyticus.";
RL Agric. Biol. Chem. 54:3227-3233 (1990).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC ATCGAT, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE BANIII ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X83417; CAA58443.1; ALT_SEQ.
CC FIR; JH0224; JH0224.
CC REBASE; 3296; M.BanIII.
CC InterPro; IPR002296; N12N6_mtfrase.
CC InterPro; IPR002052; N6_Mtase.
CC InterPro; IPR000051; SAM_bind.
CC PRINTS; PR00507; N12N6MTFRASE.
CC PROSITE; PS00092; N6_MTASE; 1.
CC TRANSFERASE; Methyltransferase; Restriction system.
CC KW SEQUENCE 580 AA; 66344 MW; 0E4DF7357EC394C7 CRC64;

Query Match 5.4%; Score 134; DB 1; Length 580;
Best Local Similarity 22.9%; Pred. No. 0.03;
Matches 75; Conservative 54; Mismatches 114; Indels 84; Gaps 18;
```

```

QY 23 QAVLGQFTPMKAATLMASMLRVDDL---GT---VRVLDPGAGVGSILTAALVDRLHTE 75
DB 12 QKETGAHYTPTDLGDIIAKRL-INELKKSGISGTTKIRGLDPSGCGELLISL-NRIAKF 69
QY 76 RPDVAHVAVETDPFVVPYLRATLEEC---RNAYGIS-YDLVEGDYLLNQAKLDG--- 128
DB 70 NNIDNIELIGIDED-----KBAIKEADFRLNEMGINDAKLTAGDFL---DMVDLEGNLS 120
QY 129 -----PFDLVIANPPYGGK---LASDSLARLATTARVAVDVPNVVAVFWRAVISLK 175
DB 121 LFDDDLKSLKIEPVDLIANPPVVRQVLGADRAQKLAFLNKGKRVLDLTHAFVAMTLQK 180
QY 176 EQGR-GVFIVPRSWAN--GPYYRQFRHLMTAVSLDLHVFESRTKVPADTKVQENVI 232
DB 181 PGGIGLIGITSNKYLANSGESIRQF-----LAENYDIIEIMD-----LGDIKLI----- 223
QY 233 AFSVRPQSSSVVLRSVAHGESSTASSVPFSA-----LVHDEDD-----DKIVHF 277
DB 224 -----FSAVLQALIFFGGRKKNGKIRQAPANFYKIYEBTDPKTEVSIRKFTLFLG 275
QY 278 AESASVPSAARFTLADLIGIGYSTGKV 304
DB 276 LESS---NTGVFNDEKFEYSVSCGLI 299

RESULT 13
TIMP ECOLI
ID _TIMP_ECOLI STANDARD; PRT; 520 AA.
AC Q47163;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Type I restriction enzyme EcopRI M protein (EC 2.1.1.72) (M.EcopRI).
GN HSDM.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CTR5X;
RX MEDLINE=94194504; PubMed=8145241;
RT Tyndall C., Meister J., Bickle T.A.;
RA "The Escherichia coli prr region encodes a functional type IC DNA
RT restriction system closely integrated with an anticodon nuclease
RL J. Mol. Biol. 237:266-274 (1994).
CC -!- FUNCTION: THE M AND S SUBUNITS TOGETHER FORM A METHYLTRANSFERASE
CC (MTASE) THAT METHYLATES TWO ADENINE RESIDUES IN COMPLEMENTARY
CC STRANDS OF BIPARTITE DNA RECOGNITION SEQUENCE. IN THE PRESENCE OF
CC THE R SUBUNIT THE COMPLEX CAN ALSO ACT AS AN ENDONUCLEASE. BINDING
CC TO THE SAME TARGET SEQUENCE BUT CUTTING THE DNA SOME DISTANCE FROM
CC THIS SITE. WHETHER THE DNA IS CUT OR MODIFIED DEPENDS ON THE
CC METHYLATION STATE OF THE TARGET SEQUENCE. WHEN THE TARGET SITE IS
CC UNMODIFIED, THE DNA IS CUT. WHEN THE TARGET SITE IS
CC HEMIMETHYLATED, THE COMPLEX ACTS AS A MAINTENANCE MTASE MODIFYING
CC THE DNA SO THAT BOTH STRANDS BECOME METHYLATED. THE ECOPRI ENZYME
CC RECOGNIZES 5'-CCA(N7)ATCG-3'.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -!- SUBUNIT: The type I restriction/modification system is composed
CC of three polypeptides R, M and S.
CC -!- MISCELLANEOUS: TYPE I RESTRICTION AND MODIFICATION ENZYMES ARE
CC COMPLEX. MULTIFUNCTIONAL SYSTEMS WHICH REQUIRE ATP, S-ADENOSYL
CC METHIONINE AND MG(2+) AS CO-FACTORS AND, IN ADDITION TO THEIR
CC ENDONUCLEOLYTIC AND METHYLASE ACTIVITIES, ARE POTENT DNA-DEPENDENT
CC ATPASES.
CC -!- SIMILARITY: WITH TYPE II RESTRICTION SYSTEM ADENINE METHYLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X75452; CRA53205.1; -.
DR REBASE; 3552; M.EcopRI.
DR InterPro; IPR004546; HsdM.
DR InterPro; IPR003665; Methylase_M.
DR InterPro; IPR003356; N6_DNA_Mtase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR Pfam; PF02506; N6_Mtase_M; 1.
DR TIGRFAMs; TIGR00497; hsdM; 1.
DR PROSITE; PS00092; N6_Mtase; 1.
DR Restriction system; Transferase; Methyltransferase.
KW SEQUENCE 520 AA; 58015 MW; 79401EC1720EB257 CRC64;

Query Match 5.3%; Score 132; DB 1; Length 520;
Best Local Similarity 21.8%; Pred. No. 0.036;
Matches 67; Conservative 48; Mismatches 121; Indels 72; Gaps 13;

QY 27 GQFTPMKAATLMASMLRVDDLRTV---RVLDPGAGVGSILTAALVDRLHTEPDVAHV 83
DB 197 GEFTPQHVSFKLIAQLA---MHGQTHVNKLYDPAAGSSL-----LLOAKQFDDHI 245
QY 84 VAVETDFVVPYLRATLEECRN---AYGISYDLVE---GDYLLNQAKLDGPPDLVIANP 137
DB 246 --IIEGFFGQGEINHTTYNLARMNMFNLINVDKFDIKLGNLTLPHPFRDEKPFDAIVSNP 303
QY 138 PYGK--LASDLSA-----RLATTARVAVDVPNVVAVFWRAVISLKEQGRGVFI-VPRSWA 189
DB 304 PYSVKWIGSDPTLINDERFAPAGVLAPKSKAOFAPVILHNLNLSAKGRAIIVCFPGIFY 363
QY 190 NGPYRQFRHLM-----TAVSLDLHVFESRTKVFADTKVK----- 226
DB 364 RGGAEQKIRQYLVNNDNNVETVISLAPNLPFGTTIANVILVLSKHT-----DTKVQFIDAS 419
QY 227 -----QENVIVAFSVRPOSSSVLSRSVAHGESIA-----SSVPFSALVHDEDD 271
DB 420 ELFKETNNILTAHIEQIMQVFASKEDVAHLAKSAVFETTVVANDYNLSVSSVVEAKDT 479
QY 272 DKIVHFAE 279
DB 480 REIIDIAE 487

RESULT 14
YY02 METJA
ID YY02_METJA STANDARD; PRT; 1181 AA.
AC Q60301;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJEC502.
GN MJEC502.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 47.4586 Seconds  
(without alignments)  
3231.066 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHP1ASTETRRQAALGKLDP.....QETLDVAVARLFCBPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTEMBL 25: \*  
2: sp\_archaea: \*  
3: sp\_bacteria: \*  
4: sp\_fungi: \*  
5: sp\_human: \*  
6: sp\_invertebrate: \*  
7: sp\_mammal: \*  
8: sp\_mmc: \*  
9: sp\_organelle: \*  
10: sp\_phase: \*  
11: sp\_plant: \*  
12: sp\_rodent: \*  
13: sp\_virus: \*  
14: sp\_vertebrate: \*  
15: sp\_unclassified: \*  
16: sp\_rvrius: \*  
17: sp\_bacteriaph: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	669.5	27.1	826	2 Q8VLS0	Q8VLS0 yersinia en
2	637	25.8	472	2 Q9LA86	Q9LA86 aeromonas h
3	379	15.4	546	16 Q8YQW1	Q8YQW1 anabaena sp
4	286	11.6	596	2 Q9RBJ7	Q9RBJ7 xanthomonas
5	246.5	10.0	600	2 P72289	P72289 rhizobium l
6	243.5	9.9	236	16 Q8YQW2	Q8YQW2 xanthomonas
7	179.5	7.3	428	2 Q56238	Q56238 thermus the
8	176.5	7.1	543	16 Q8F7P1	Q8F7P1 leptospira
9	176	7.1	368	12 Q37396	Q37396 chlorella v
10	175	7.1	848	2 Q8VVK9	Q8VVK9 corynebacte
11	174.5	7.1	546	2 Q9KVZ8	Q9KVZ8 xanthomonas
12	174.5	7.1	571	2 Q8VT54	Q8VT54 xanthomonas
13	172.5	7.0	862	2 Q9EXA5	Q9EXA5 neisseria m
14	165	6.7	487	16 Q25210	Q25210 helicobacte
15	163	6.6	540	2 Q53293	Q53293 acinetobact
16	161	6.5	543	16 Q9ZM08	Q9ZM08 helicobacte

17 159.5 6.5 435 2 Q93HR2  
18 159 6.4 528 16 Q8PJU4  
19 158.5 6.4 541 2 Q56752  
20 157 6.4 379 2 Q9KGT6  
21 147.5 6.0 568 2 Q9KJH2  
22 147 6.0 379 16 Q25568  
23 147 6.0 579 2 Q9RQK2  
24 141 5.7 571 16 Q97DB4  
25 141 5.7 832 16 Q8UGB6  
26 140 5.7 555 16 Q9CCK4  
27 137 5.5 514 16 Q9K001  
28 136 5.5 684 16 Q8EGV1  
29 135 5.5 553 16 P96868  
30 135 5.5 553 16 Q7TWV2  
31 135 5.5 2162 2 Q9FB27  
32 134 5.4 514 16 Q9JVI6  
33 134 5.4 534 16 Q8DUM7  
34 134 5.4 800 2 Q93C65  
35 134 5.4 1105 2 Q8GJ86  
36 133.5 5.4 381 16 Q9ZKT9  
37 132 5.3 357 12 Q96719  
38 131.5 5.3 620 2 Q56788  
39 131 5.3 1061 2 Q8RNV3  
40 130.5 5.3 369 16 Q8PKD3  
41 130 5.3 526 16 Q92S33  
42 130 5.3 545 16 Q25223  
43 127.5 5.2 352 10 Q8L867  
44 127.5 5.2 380 10 Q9FK02  
45 126.5 5.1 808 17 Q8PUN8

#### ALIGNMENTS

RESULT 1  
Q8VLS0  
ID Q8VLS0 PRELIMINARY; PRT; 826 AA.  
AC Q8VLS0; 2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Methyltransferase-endonuclease.  
GN YENI  
OS Yersinia enterocolitica.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON=IS1222;  
RA Rakin A.V.;  
RT "Characterization of the restriction-modification system of Y.  
RT enterocolitica biotype 1B";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ414030; CAC95150.1; -  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004519; F:endonuclease activity; IEA.  
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.  
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0006306; P:DNA methylation; IEA.  
DR InterPro; IPR002296; N12N6\_mtfase.  
DR InterPro; IPR002052; N6\_Mtase.  
DR InterPro; IPR000051; SAM bind.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
KW Transferase; Endonuclease.  
SQ SEQUENCE 826 AA; 93348 MW; 148432FD8E798843 CRC64;

Query Match 27.1%; Score 669.5; DB 2; Length 826;  
Best Local Similarity 32.5%; Pred No. 2.1e-42;  
Matches 162; Conservative 107; Mismatches 186; Indels 43; Gaps 13;





```
Db 60 E-TVPELLYPAIKTKVLP--PKKPKRSIAIVQNETDKWLIPSGWVLTFRSAKEE 115
QY 377 RRRVIAAVMD--GASRVLDNKNYLVHESQRPLEKNVARGMLNLSNTVLDQYFRAFSGH 434
Db 116 KRRVVAACSPVDAPVLGIENHLNYYHSGQGMNPDLAGLAFLSSILLDSYFRFSGH 175
QY 435 TQVNAAGDRLRPLFCREDLILIAKVVPDGLPDQETLDAVAVARLFCPEIPES 484
Db 176 TQVNAATDLRIKYPCKDDILKLSGIGDSCLDQALDQVHTVHKTLSIMSEA 225

RESULT 4
Q9RBJ7
ID Q9RBJ7 PRELIMINARY; PRT; 596 AA.
AC Q9RBJ7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyltransferase homolog M.XphI.
GN XPHIM.
OS Xanthomonas campestris (pv. phaseoli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxID=29445;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xcp 73;
RA Lai J.Y., Yang M.T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF042157; AAP22366.1; -.
DR HSSP; P14385; 2ADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mfrase.
DR InterPro; IPR002052; N6_Mcase.
DR PRINTS; PR00507; N12NGMTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase.
SQ SEQUENCE 596 AA; 64745 MW; 87AED38BLC2995F3 CRC64;

Query Match 11.6%; Score 286; DB 2; Length 596;
Best Local Similarity 26.5%; Pred. No. 3.2e-13;
Matches 132; Conservative 55; Mismatches 186; Indels 126; Gaps 20;

QY 27 GQFFTPMKAATLMASMLRVDDL-R-GTVRVLDPGAGVSLTAALVDRLHTE-----RDPVAV 81
Db 125 GVETTPALAEILLDQAAAGLQWSRAHVLDPDPAAGAGFLVPAARLLKSLGDCSPAVAL 184
QY 82 HVAV-----ETDPF-----VVPYL-----RATLEECRNAYGISYDLVEGD 117
Db 185 HNLSARLRGFLDPFAAMWAQVFEAAALPLIVACGRPTAVLTVDG-----LSITKSG- 240
QY 118 YLLNQAKLDGPDLLVIANPPYKGLASDSLARIATTARAVDVPNVAVFVRVAVISLKEQ 177
Db 241 -----FDLVGNPPFGRLKLAERREYFSSRLYGHANLYGLFMDLAVLAKPD 288
QY 178 GRGVFIVPSWANGPYRQFR---HMLMTAVSLDLHVFESRTKVPFADTKVKQENVIVAF 234
Db 289 GLVSFLTPSSFLAGEYFKMLRAVLHKEAPPVSLDFV---TARKGVFDD---VLQETVLATY 343
QY 235 S-----VRPQSSSVLRSVAHGEESIASVPFSALVHDEDDDKIVHFEASAS 282
Db 344 RXGAKRARAVVSFIEAQPGVPVKAEPAGHTLPRKATAPWFLPRHADE-----AELAK 396
QY 283 VPSAARTLADLIGIVSTGKVDVFRNQYLTNDLSDAGVVPVYQSNIRS-GKI----- 335
Db 397 RLRAKSAPLADMGYKVTGTLVWNRKPKQLCDSEEA-GTVPLVWAESVTSDBGFVLRAEK 455
QY 336 -----DWPQVGARKPGQFVAVEDVALRQLLPQSGSVVVKVKTAKEDRRRVIA 382
Db 456 RNHKPFLRLOPGDDMLV--RKP-----CVLIQRTTAKQARLLIA 494
```

```
QY 383 AVWDA-----SRVALDNKNYLVHESQRPLEKNVARG---LMLNLSNTVLDQYFRAFSGH 434
Db 495 AEMPAKFIRKHAGVVTIENHLM-----IPTVENPAVSPALLAAFLNSDAADRAFCMSG 550
QY 435 TQVNAAGDRLRPLFCREDL 453
Db 551 VAVSAVELENPLPTASDL 569

RESULT 5
P72289
ID P72289 PRELIMINARY; PRT; 600 AA.
AC P72289;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methylase.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OC NCBI_TaxID=384;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98055154; PubMed=9393436;
RA Rochepeau P., Selinger L.B., Hynes M.F.;
RT "Transposon-like structure of a new plasmid encoded restriction-
modification system in Rhizobium leguminosarum VF39SM.";
RL Mol. Gen. Genet. 256:387-396(1997).
DR EMBL; X99520; CAA67873.1; -.
DR HSSP; P14385; 2ADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mfrase.
DR InterPro; IPR002052; N6_Mcase.
DR PRINTS; PR00507; N12NGMTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase.
SQ SEQUENCE 600 AA; 65707 MW; C3909584B71D69DF CRC64;

Query Match 10.0%; Score 246.5; DB 2; Length 600;
Best Local Similarity 25.8%; Pred. No. 3.4e-10;
Matches 137; Conservative 70; Mismatches 230; Indels 95; Gaps 25;

QY 3 PIASTETROAALGKLDL-TTQAVLGQFFTPMKAATLMASMLRVDDL-RG-----TVRVLD 57
Db 104 PIESAFLSACVTAMLPQDIRSEWGAFTY-----PALTKLMELAQEAGIDWRAARVLD 160
QY 58 GAGVSLTAALVDRLHT-----ERPD-VAVVAVAVETDPFVFPYLRATLEECRNAY 108
Db 161 ACGGAFLLPVALRMQOALQALSPGELLDFAGHLSGDFIDPFAAQLTQTWLE---IAFA 217
QY 109 ISYDLVEG---DYLLNQAKLDG-----FDLVIANPPYKGLASDSLARIATTARAVDP 160
Db 218 -SLSMOTGTPFFPAFVRKVCNSLEQPVSSKRFDLVIGNPPYGRVLRNARLRERYRSLFGHA 276
QY 161 NVYVAFVRAVTSLEKQGRGVFTVPSRWANGPYRQFRHMLMTAVSLDLHVFESRTKVP 220
Db 277 NMVGLFTDLQWARKGVAVVPTGFLAGEYFKALRALAKADAPFAIDFITERRGIF 336
QY 221 AOTKVKQENVIVAF-----SVRFQSSSVLRSVAHGEES-----IASVPFS 263
Db 337 DD--VLQEAALLATYRRGGSLGSPVTVHYFSVNGTAQVTHAGEFHLPKDASQPLARVP-- 392
QY 264 ALVHDEDDDKIVHFEASASVPSAARTLADLIGIVSTGKV-----DFNRQYLTNDL 317
Db 393 -----DDGVLY--KQLSRLP-----RLKDWGKVTGTPLVWNRKPKQPRGA----- 433
QY 318 ASGVPMVYQSNIRS-GKIDM-POVGARKPGQFVAVEDVALRQLLPQSGSVVVKVKTAK 375
Db 434 GKDTFVIAEAVSADGKFSFAEKRNQHPFKVRPEDQWLEVSVP-----CVLIQRTTSKE 490
```

```

Qy 376 DRRVIAAVWDG-----SRVALDNKTNLYHESQ-RPLEKNVARGMLMLNSTVLDOYFR 429
Db 491 QARLIIAAELPEAFIKAHGRVIVENHLNMVPTAGKPRVSTAVVAV--LNSKIADRAFR 548
Qy 430 AFSGHTOVNAGDLRLPLFLCKREDIILLAKVVPDGL--PDQETILDVAVARLF 478
Db 549 CIGSVAVSAFEALPLPKPEAL---KAVGDLILAKHADQAVDAALERLF 596

RESULT 6
Q8YQW2 PRELIMINARY; PRT; 236 AA.
AC Q8YQW2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Site-specific DNA-methyltransferase.
GN ALB3700.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:203-213 (2001).
DR EMBL; AP003594; BAB75399.1; -.
DR PIR; A52268; AE2268.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; F:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 236 AA; 26734 MW; 63C1EC76516FF9ED CRC64;

Query Match 9.9%; Score 243.5; DB 16; Length 236;
Best Local Similarity 31.3%; Pred. No. 1.4e-10; Indels 17; Gaps 6;
Matches 68; Conservative 41; Mismatches 91;

Qy 6 STETTRQAALGKLDPTTQAVLGQFFTPMKAATLMASMLRVDDLRGTVRVLDPCAGVGSILT 65
Db 8 STDINRINFSSNLNLRRLRGELGFTLPAPLRFVVG--QFNLSLGHISLIDPCAGVGSILT 65
Qy 66 AALVDRLHTERPDV-AHVAVAVETDPFVVPYRATLEECRNAY--GI--SYDLVEGDYL 119
Db 66 AAFVERLLANSHEVKSCFTVVEVEFIFLPLKQCLIDCCTALENGIAANYCLYEKNF- 124
Qy 120 LNOGAKLDGP-----PDLVTANPPYKGLASDSLRLATTARAVDVENVYVAVWRAV 171
Db 125 IDANSINPLSTSAINFTHALINPLYKKNKNSIEKKIISQMGITVNLISAFVWLII 184
Qy 172 ISLKEQGRGVFIPRSWANGPYRQFRHMLMTAVSLD 208
Db 185 IRLVEGEIVAITPRFCNGSYRFRKFSLEQMKLE 221

RESULT 7
Q56238 PRELIMINARY; PRT; 428 AA.
AC Q56238;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methylase.

```

```

GN M. TTHB81.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=92201699; PubMed=1339363;
RA Brany F., Danzitz M., Zebala J., Mayer A.;
RT "Cloning and sequencing of genes encoding the TthB81 restriction and
RT modification enzymes: comparison with the isoschizomeric TaqI
RT enzymes.";
RL Gene 112:31-95 (1992).
RN [2]
RP SEQUENCE FROM N.A.
EX MEDLINE=92201707; PubMed=1551602;
RA Barany F., Siatko B., Danzitz M., Cowburn D., Schildkraut I.,
RA Wilson G.G.;
RT "The corrected nucleotide sequences of the TaqI restriction and
RT modification enzymes reveal a thirteen-codon overlap.";
RL Gene 112:91-95 (1992).
DR EMBL; M76882; AAA27491.1; -.
DR HSP; P14385; 2ADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 428 AA; 48478 MW; F24945880651D9CD CRC64;

Query Match 7.3%; Score 179.5; DB 2; Length 428;
Best Local Similarity 22.4%; Pred. No. 2.8e-05; Indels 153; Gaps 22;
Matches 110; Conservative 60; Mismatches 168;

Qy 18 LDPTTQA-VLGOFFTPMKAATLMASMLRVDDLRGTVRVLDPCAGVGSILTALVDRLHTER 76
Db 7 LDPTASGRSLGRVETP---PGLVRFMVGLAEARKGVRLPEACADGPFL-----RAFREA 58
Qy 77 PDVAVHVAVETDPFVW---PYLRATLEECRNAYGISYDLVEGDYLLNOGAKLDGDFDLV 133
Db 59 HGTGYRFVGVGEIDPHALDLPWAEG-----VVADFLLEPGE---AFDLI 100
Qy 134 IANPPYKGLASDSLRL-----ATTARAVDPNVYVAVFWRAVISLKEQGRGVF 182
Db 101 LGNPPYGIYGEASKYPIHVLREVKGLYKKTLSWKGYKNLYGAFIEKSVLLREGTLVF 160
Qy 183 IYPRSWANGPYRQFRHMLMTAVSLDILH---VFESRTKVA-----DTK 224
Db 161 VVPATWLVDLSLSLRSFLAREGRIEYVYLGEPGR-KVSAVLRFRKGKGLALWDTR 219
Qy 225 VKQENVI-VAFSVRPOSSSVIISRSVAHGEESIASVPFSALVHDEDDDKIVHFAFSASV 283
Db 220 RDGETFTPLWSEKPEWKGEIIRFTGWTREMEASGPPLGSLFH-----I 264
Qy 284 PSNAFTLADLIGVSTGKVDPRNEQYLTDNLDSGVVPMVYQSNIRSGKIDW----- 337
Db 265 PFAARSP-----EFKHPAVOKPEP-FGLVPLVTGRNLKPGWIDYESNHS 309
Qy 338 ---POVGARKPGFVAVEDVALRQLLPQGSYVVKRQTAKEDRRRRVIAAVWDGAS----- 389
Db 310 LMPKERAKELRDFYATPHL-----VVAHTKGTG-----VVAANDRAYPWRE 352
Qy 390 -----RVALDNKTNLYHESORPLEKNVARGMLMLNSTVLDOY-----FRFSGHTQV 437
Db 353 EFHLLPKGVGELD-----PL-----FLVEMLNSDKIQEVYKVTLYRDFVPHLTL 395
Qy 438 NAGDLRLPL 448

```



```

OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LP-6;
RA Tauch A., Goetker S., Puehler A., Kalinowski J., Thierbach G.;
RT "Antimicrobial resistance in Corynebacterium glutamicum: The 27.8-kb
RT R-plasmid pRST3 encodes the aminoglycoside adenylyltransferase gene
RT cassette aadA8 and the regulated tetracycline efflux system Tet 33.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ420072; CADI2208.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Plasmid.
SQ SEQUENCE 848 AA; 95859 MW; 879C7E07C31E0613 CRC64;

Query Match 7.1%; Score 175; DB 2; Length 848;
Best Local Similarity 20.0%; Pred. No. 0.00018;
Matches 110; Conservative 80; Mismatches 173; Indels 188; Gaps 23;

QY 5 ASTETTRQAALGKLDPTTQAVLGQFFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVGS 64
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 294 AKTETIQIFMSFPAVEFKELSYFTPI---SLIETVEMVDIGITDKVDPDPMGTGDP 350
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 65 TAAVLDRLHTRERPDVAHV--VVAVETDPFVVPYLRATLEECRNAYGISYDLVEGDYLLNQ 122
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 351 ---LVEALEKRGDDDIHQRLFAGDRDP-----SAYELAIVNMLNK 389
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 123 GAKLDGPPDLV-----IANPPYG-----KLADS 146
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 390 ---DQGTGLVLQDSIKNHTLWANEMNVALCNPPFGSRTVERSKSVLEAYDLGYKWEEDS 445
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 147 LARLATTARADVNPVYAFVRAVISLKEQGVFIIPRSWANGPYVQRRHMLMTAVS 206
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 446 NGVMYKTDVLSQQGLGILFIERCKWMLTDQRLGILPEGYSKAKYILKQWIIDHFI 505
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 207 LDILHVFESRTKVFADTKVQENIVAFSVRQSSSVLSRSVAHGEESIASV----- 260
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 506 VHAVVELPRMFKSDADLRSMILLLEKSDASRNA-----GRKIYASWVRKVGYK 556
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 261 ---PFSA-----LVHDEDDDKIV-----HFABSAS 282
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 557 LAGDFSATPQDPETGLVLHDDENEPFLDSDFNRVLEEKQKQVTDREWEAGHLKDVLS 616
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 283 VPSAARFTLADLGIGVSTGKVDFRNROYLTD-NLDAS---GVVPMVYQSNIR----- 331
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 617 HPE-----LDLKPRLVP-RAIQUMTSINDPSVVRVLDVAEIVTDNIRVASPEN 665
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 332 SGKIDWPQVG--ARKPQGVAVEDVALRQLLPQGSYVVVVKRQTA-----K 374
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 666 RSKLPFPIEGQINRAVEGIVSPCD-----PERGWHIAQRKTSNCPVEPQPDILVGLVR 718
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 375 EDRRRY-----IAAVWDGASRVADLNK---TNYLHESQRPLEKNVARGMLMLNS- 421
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 719 PERRNVGLVDSEGIFVGKIDGVAVVRASNKDPFGWLFYALRRETRI-----QMWIKAG 774
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 422 -----TVLDQ 426
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 775 GTSYGLKLDLQ 785
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q9KVZ8 PRELIMINARY; PRT; 546 AA.
ID Q9KVZ8;
AC Q9KVZ8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE NCBI_TaxID=133782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LP-6;
RA Morita M., Sugino Y.;
RT "Nucleotide sequences of the XhoI methylase and endonuclease genes.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB046567; BAB03596.1; -.
DR HSSP; P14385; 2ADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Methyltransferase.
SQ SEQUENCE 546 AA; 61025 MW; D9EF9B9D4E137BD2 CRC64;

Query Match 7.1%; Score 174.5; DB 2; Length 546;
Best Local Similarity 23.8%; Pred. No. 9.9e-05;
Matches 100; Conservative 63; Mismatches 172; Indels 85; Gaps 21;

QY 27 GQFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVGSLSAALVDRL----- 72
Db 6 GAIFTRREVVEFILLDVGTYEDRLAQTLLKLEPSAGHADFLPLIIGRLVRSYVAHGDLT 65

QY 73 ---HTERPDV---AVHVVAVETDPV---PYRATLEECRNAYGISYD-LVEGDYLLAQ 123
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 66 RGAQTLGPAIRAYEVHESSELTARSVVVAELRLGVKKA--ADGLRTWLVRADFLM--- 121
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 124 AKLDGPPDLVIANPPYGK---LASDLARLATTARAV-DVPNVYVAFVRAVISLKEQGR 179
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 122 APLPHTFDVGNPPYVQELIPAVLLARYARFKTLDRLADLYPFYERCLDLAPGR 181
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 180 GVFIIPRSWANGPYVQRRHMLMTAVSLDILHVFESRTKVFADTKVQENIV--AFSV- 236
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 182 LGFICTDRWTKNKGGLR--AMVSEEFSLTFVD-----LVDTQAFLSNVMTYPAITVI 234
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 237 -RQSSSVLSRSVAHGEESIASVPPFSAL-----VHDEDDDKIV 275
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 235 ERPKPKSKARPRVAY-RPAISAEV-FGPLAKAMTGTKLNHKAGVVMGVSNGSEPWIL 292
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 276 HPAESASVPSAARFTLADI-----GIGVSTGKVDFRNROYLTD---NLDASGVWPM 324
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 293 HQADRLALVRLLEETLPTLEEAGCKVGIVATG-----NDGVIGDMKTLNVEPSKLP 347
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 325 VYQSNIRSGKIDWPQVGARKP-QGFVAVEDVA---LRQLLPQGSYVVVVKRQTAKEDE 379
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 348 ARTQDLRGSIDWQGVNLNPFEDGQVVDLASYPKFAAYLQEHAIQIKARHVAKNPER 407
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
Q8VT54 PRELIMINARY; PRT; 571 AA.
ID Q8VT54;
AC Q8VT54;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE XhoI modification methylase.
GN XHOIM.
OS Xanthomonas campestris pv. holcicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=133782;
RN [1]
RP SEQUENCE FROM N.A.
RA Anton B.P., Benner J.S., Moran L.S., Morita M., Sugino Y.,
RA Brooks J.E.;
RT "The XhoI restriction-modification system is part of an isochizomeric
```



Search completed: October 1, 2004, 16:10:27  
Job time : 51.4586 secs

```

RESULT 15
053293
ID Q53293 PRELIMINARY; PRT; 540 AA.
AC Q53293;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Methyltransferase.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=91345839; PubMed=1368703;
RA Kawakami B., Christophe H., Nagatomo M., Oka M.;
RT "Cloning and nucleotide sequences of the AclI restriction-modification
genes in Acinetobacter calcoaceticus.";
RL Agric. Biol. Chem. 55:1553-1559(1991).
DR EMBL: S50843; AAC60387.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008470; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro: IPR002296; N12N6 mtrase.

```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 63.6786 Seconds  
(without alignments)  
2156.424 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHPASTETRRQAALGKLDP.....QETLDVAVVRLFCBPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseqp29Jan04.\*  
2: Geneseqp1980s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	223	9.0	1068	6	ABR43481	Abm67143 Photorhab
2	174	7.0	550	6	ABM67143	Abm67143 Photorhab
3	164	6.6	543	2	AAW20785	Aaw20785 H. pylori
4	151	6.1	533	6	ABP79018	Abp79018 N. gonorr
5	147	6.0	626	3	AAV88485	Aay88485 DraIII me
6	141	5.7	571	6	ABU23716	Abu23716 Protein e
7	137	5.5	514	3	AAV75218	Aay75218 Neisseria
8	137	5.5	519	6	ABM70049	Abm70049 Photorhab
9	135	5.5	510	6	ABP78365	Abp78365 N. gonorr
10	135	5.5	514	3	AAV75217	Aay75217 Neisseria
11	135	5.5	514	3	AAV75217	Aay75217 Neisseria
12	135	5.5	2162	3	AAU07560	Abu37303 Protein e
13	134	5.4	514	3	AAV75219	Aay75219 Neisseria
14	134	5.4	514	6	ABU37930	Abu37930 Protein e
15	134	5.4	534	6	ABU44389	Abu44389 Protein e
16	130	5.3	545	2	AAW98867	Aaw98867 H. pylori
17	126	5.1	502	2	AAW98867	Aaw98867 H. pylori
18	125.5	5.1	317	4	AAW98867	Aaw98867 H. pylori
19	125.5	5.1	317	6	AAW98867	Aaw98867 H. pylori
20	125.5	5.1	317	6	AAW98867	Aaw98867 H. pylori
21	124.5	5.0	295	6	ABU36560	Abu36560 Protein e
22	124.5	5.0	511	4	AAU25782	Abu25782 Protein e
23	124.5	5.0	530	6	ABU28930	Abu28930 Protein e
24	122	4.9	358	2	AAW98649	Aaw98649 H. pylori
25	122	4.9	530	6	ABU29993	Abu29993 Protein e

## ALIGNMENTS

## RESULT 1

ABR43481

ID ABR43481 standard; protein; 1068 AA.

XX AC ABR43481;

XX 23-OCT-2003 (revised)

DT 16-JUL-2003 (first entry)

XX DE Bacillus stearothermophilus B61 methylase protein SEQ ID NO:11.

XX KW Bacillus stearothermophilus B61; BsmI; methylase; bsmBIM; enzyme; molecular cloning; gene characterisation.

XX OS Geobacillus stearothermophilus.

XX PN EP1298212-A2.

XX PD 02-APR-2003.

XX PF 27-SEP-2002; 2002EP-00256756.

XX PR 28-SEP-2001; 2001US-00966997.

XX PA (NEW ) NEW ENGLAND BIOLABS INC.

XX PI Xu S, Dore A, Hume A, Pellitier J, Zhou J;

XX DR WPI; 2003-395598/38.

XX DR N-PSDB; ACC69532.

XX PT New DNA segment coding for the BsmBI restriction endonuclease and/or BsmBI methylase, useful for molecular cloning and gene characterization, and in producing restriction endonucleases and modification methylases.

XX PS Example 1; Fig 2A-D; 35pp; English.

XX CC The present sequence represents Bacillus stearothermophilus B61 BsmBI methylase (bsmBIM). The present invention also describes BsmBI restriction endonuclease (bsmBIR) isolated from B. stearothermophilus B61. BsmBIR and bsmBIM can be isolated from a DNA segment obtainable from American Type Culture Collection (ATCC) Accession Number PTA-3739. The DNA segment is useful in molecular cloning and gene characterisation, and in producing BsmBI restriction endonucleases and modification methylases. (Updated on 23-OCT-2003 to standardise OS field)

XX SQ Sequence 1068 AA;

Adc96965 E. faeciu  
Aay04661 L.lactis  
Abp25613 Streptoco  
Adc96966 E. faeciu  
Abu48789 Protein e  
Aaw98507 H. pylori  
Abu02489 S. pneumo  
Abu44324 Protein e  
Abb54003 Lactococ  
Abu16328 Protein e  
Aay81754 Streptoco  
Aau38061 Streptoco  
Abu02552 S. pneumo  
Abu46269 Protein e  
Abu20637 Protein e  
Abu30617 Protein e  
Abu16331 Protein e  
Abm71646 Staphyloc  
Aau37589 Streptoco  
Abu46253 Protein e



Query Match 9.0%; Score 223; DB 6; Length 1068;  
 Best Local Similarity 20.7%; Pred. No. 3.8e-13;  
 Matches 112; Conservative 86; Mismatches 219; Indels 124; Gaps 21;

QY 21 TTQAVLGQFFTPMKAATLMAFMLRVLDLGRVTVRLDPGAGVGSUTA-ALVDRHLHTERPDV 79  
 Db 118 TASKAYGYFFTPISLGTWVKALKDKPKNLKSTVDPAACGIGSLALALI-----YNPEI 172

QY 80 AVHVAVVETDPFVVPYLRATLEECRNAYGIS--VDLVEG-----DYLNOGAKLDGPFDLVI 134  
 Db 173 E-NVVGIELDSFTANISHKLLVRISKDLGITPKIKIINQNFIDVLYNVEEHEKFDLLI 231

QY 135 ANPYGKLI-----ASDLSARLATTARAVDVP 160  
 Db 232 MNPPYGRVRLKSLTNKTKSGLTGISELEKLRRETIINADLRKFKFASVGLGKTP 291

QY 161 NVYAFVWVRVVISLKEQGRGVFIVPRSWANGPYVRQFRHMLTAVSLDILHVFESRTKVF 220  
 Db 292 EYSKVFLAISTKIYKQNGYVIAITPSSWLGDESGRELKYLVENHGISCIWNFKESAKLF 351

QY 221 ADTKVKQENVIAFVRPQSSVLSRVAHGEESIASVPSALVHDEDDDKIVHFAES 280  
 Db 352 SG--VNQPTTVKIKVNSNEKIEIQG-----PLSSL---EELGRDIQYLDI 393

QY 281 ASVP--SAARFTLADLGIG-----VSTGKVDPRNQYLTDLNLDASGVVPM 324  
 Db 394 CNIKYSPEWYRPOCGNERAKILSKLNHAPLSHKKIYINLRGELDTSHKLLSDNPN 453

QY 325 VQOSNRSKIDWPVGARKPO-----GFVAVEDVAL-----ROLIIPQ 362  
 Db 454 HWRL-IRGDHVE--KENLKNPSESEKLGIV---DHQLFKRMGKSNKLRHKNWRTILPQ 507

QY 363 GSYVVVXQRTAKEDRRRVIAAVWDGASRVALDNKNYLYH-ESQRPLEKXVARGMLWLS 421  
 Db 508 CSYM-----NKKRIEACIVEPNNIIA--NSCNYITLEDNCNELVDNLLL-LCAIINS 556

QY 422 TVLDQYFRAFSGHTQVWAGDLRLPLC-----REDLILAKVVPDGLPQETLDVAVARL 477  
 Db 557 AVIEWREFLNSNNHVSNYEIDFPIKFDTEMTMLKGFHFKPIENWKSIEALIALM 616

QY 478 F 478  
 Db 617 Y 617

RESULT 2  
 ID ABM67143  
 XX ABM67143 standard; protein; 550 AA.  
 AC ABM67143;  
 XX 20-NOV-2003 (first entry)  
 DE Photorhabdus luminescens protein sequence #240.  
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 XX whooping cough.  
 OS Photorhabdus luminescens.  
 XX WO200294867-A2.  
 XX 28-NOV-2002.  
 XX 07-FEB-2002; 2002WO-IB003040.  
 PF 07-FEB-2001; 2001FR-00001659.  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX

PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 DR WPI; 2003-148459/14.  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 PT Claim 2; SEQ ID NO 240; 1205pp; French.  
 XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX Sequence 550 AA;

Query Match 7.0%; Score 174; DB 6; Length 550;  
 Best Local Similarity 21.7%; Pred. No. 1.6e-08;  
 Matches 106; Conservative 71; Mismatches 176; Indels 136; Gaps 23;

QY 53 RVLDPCAGVGSITAAALVDRL-----HTERPD-----VAHVAVETDPF-----VVP 94  
 Db 39 RLLEPSFGGDFLLPIERLLSARARPSGALDELTAIRAVELHHTFKSTKAAYVT 98

QY 95 YLRATLEECRNAYGI-SYDLVEGDYLLNQAKLDGDFDLVIANPPV-----GKIASDS 146  
 Db 99 LLRRQGIANAARAVTLASYWLSQDFLL---APLEGQDFVGNPPVROELIPAPLLAEY 155

QY 147 LARLATTARADVNPVYVAFVWVRVVISLKEQGRGVFIVPRSWANGPYVRQFRHMLTAVS 206  
 Db 156 RSKYLT---MYDRADIYIIFERSLSVSDGSLGFCADRWMKNRYGGPLR-----SLV 207

QY 207 LDILHVFESRTKVF---ADTKVKQENVI-----VAFSVRQSSSVVLS-- 246  
 Db 208 AERFHL-----KVYVDMVDTFPAHSDVIAYPAITISREVPRTRIAYRPAIDPATLTAL 262

QY 247 -----RSVAHGEESTASSVPFSAVHDEDDDKIVHFAESASVPSAA 287  
 Db 263 AGLKAPSLPKVGPARELAQVTNGSE-----PW--LJESFQMAVIRLE-----G 307

QY 288 RFTLAD-----LGIGVSTGKVDPRNQYLT-----NLDASGVVPMYQSNIRSKIDWP 338  
 Db 308 DFFLLEAGCKVGIGVATG-----ADKAFIGDFDALNVEPRDKVPLVTTKDIWTEVWR 362

QY 339 QVGARKP-----QGFVAVEDV-ALRQLLPQSGYVVVVRQTAKE-----D 376  
 Db 363 GLGVINFPFQGGGLVDLDEYPRRLRYLEARREVIAGRHCAQKAPANWYRTIDRTPSLAA 422

QY 377 RRVIAAVWDGASRVALDNKNYLYHESQRPLEKN--VARGMLWLSNSTVLDQYFRAPSGH 434  
 Db 423 RPKILLIPDKQAHIVYENGELYPHHNLVYVTSDDDLRALQAVLLSAVTRLFVATYS-- 480

QY 435 TVYNAGDLR 443  
 Db 481 TKMRGGFLR 489

```

RESULT 3
AAW20785
XX AAW20785 standard; protein; 543 AA.
XX AC
XX AAW20785;
XX DT
XX 16-JUL-1997 (first entry)
XX DE
XX H. pylori cytoplasmic protein, 07gp11807orf35.
XX KW
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bacterium; life cycle; activator;
XX bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
XX diagnosis.
XX OS
XX Helicobacter pylori.
XX PN
XX WO9640893-A1.
XX PD
XX 19-DEC-1996.
XX PF
XX 06-JUN-1996; 96WO-US009122.
XX PR
XX 07-JUN-1995; 95US-00487032.
XX PR
XX 01-APR-1996; 96US-00630405.
XX XX
XX (ASTR ) ASTRA AB.
XX PA
XX Smith D, Berglindh OT, Mellgaerd BL;
XX PI
XX WPI; 1997-052306/05.
XX DR
XX N-PSDB; AAT68038.
XX XX
XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
XX useful for vaccines to treat or prevent H. pylori infection, and to
XX detect Helicobacter.
XX PS
XX Claim 61; Page 1194-95; 1481pp; English.
XX CC
XX This sequence represents a H. pylori cytoplasmic protein involved in
XX genome replication, transcription, recombination and repair. The protein
XX may be used in a vaccine to prevent or treat H. pylori infection or to
XX identify H. pylori polypeptide binding compounds, useful as potential H.
XX pylori life cycle activators or inhibitors. The genomic sequence of H.
XX pylori (ATCC 55679) was determined from overlapping contigs generated by
XX mechanically shearing the bacterial DNA. The sequences were analysed for
XX ORF of at least 180 nucleotides, and the predicted coding regions defined
XX by computer evaluation. To identify likely H. pylori antigens for vaccine
XX development, the amino acid sequences predicted from various ORF were
XX analysed for significant homology to other known or exported membrane
XX proteins. Having identified and determined the sequences of interest,
XX particular regions can be isolated from H. pylori by PCR amplification
XX for recombinant polypeptide production, e.g. in E. coli hosts
XX
XX Sequence 543 AA;
XX SQ
Query Match 6.6%; Score 164; DB 2; Length 543;
Best Local Similarity 23.3%; Pred.No.1.8e-07;
Matches 78; Conservative 61; Mismatches 124; Indels 72; Gaps 16;
QY 28 QFTPMKAATLMSMLRVDDLGRVRLVDPGAGVSLTAALVDRHLHTRPDVAHVHVE 87
DB 208 EYVTPLSIASIIAKLL-VNKPTQSVKIYDPSAGTGLLMAHAHQIGTD--SCTLYAQDIS 264
QY 88 TDFVFPVPLRATLEECRNAYGTSYDL---VEGDYLLNQAKLD--GPFDLVIANPPYGLK 142
DB 265 QKSLRMLKLNLIIND-----LTHSLRHAIEGNTLNPVHSDHKGMDFIVSNPPFKLD 318
QY 143 ASDSLARLATTAR--AVDYPN-----VYVAFWRAVISLKEQGQGVFIVR---S 187
DB 319 FSNHAEISQNKNDPFLGVFNIPKNDKSMPIYTLFFQCLNMLSPKKGAIIVPTGFTS 378

```

---

```

QY 188 WANGPYRQFRLHMLTAYSLDILHVFESRTKVADTKVKQENV-IVAFSVRQSSSVVLS 246
DB 379 AKSGVNNKVRHLVDERL---VYGVICMPSQVFANTGT---NVSIFFTQKTPSAKEVILI 432
QY 247 RSVAGGEE-----SIASSVPFSGALV-HDEDDDKIVHFAESA 281
DB 433 DASKLGEYEYTKNKKTRLPSPDMDLILETQNKRAKPSDFCALVSFDEITEK-----NY 486
QY 282 SVPSAARTLADLIGVSTGKVDPRN--ROYLTD 314
DB 487 SLNPGQYFTIEDTSETISQ---AEFENLMQOYSSE 518
RESULT 4
ABP79018
ID ABP79018 standard; protein; 533 AA.
XX AC
XX ABP79018;
XX DT
XX 07-MAR-2003 (first entry)
XX DE
XX N. gonorrhoeae amino acid sequence SEQ ID 4566.
XX KW
XX Antibacterial; infection; vaccine; gene therapy.
XX OS
XX Neisseria gonorrhoeae.
XX XX
XX WO200279243-A2.
XX PD
XX 10-OCT-2002.
XX PF
XX 12-FEB-2002; 2002WO-IB002069.
XX PR
XX 12-FEB-2001; 2001GB-00003424.
XX PA
XX (CHIR-) CHIRON SPA.
XX PI
XX Fontana MR, Pizza M, Masignani V, Monaci E;
XX DR
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ39988.
XX XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX PS
XX Disclosure; Page 523; 815pp; English.
XX CC
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid
XX molecules of the invention
XX
XX Sequence 533 AA;
XX SQ
Query Match 6.1%; Score 151; DB 6; Length 533;
Best Local Similarity 21.5%; Pred.No.4e-06;
Matches 79; Conservative 65; Mismatches 128; Indels 96; Gaps 17;
QY 28 QFTPMKAATLMSMLRVDDLGRVTR---VLDPGAGVSLTAALVDRHLHTRPDVAHVHVE 84
DB 190 EYVTPHVARIMADILVPEDVREGVRSVDYDPSAGSGLT---LMNVAH----- 235
QY 85 AVETDFVFPVPLRATLEECRNAYGTSY-----DLVEGDYLLNQAK-LDG---PF 130
DB 236 AIGEDKCM1-YTQDISQKSSNLLRLNLIINLVHSLNNVQGNITLSPAHKDASGLKXF 294
QY 131 DLVIANPPY-----GKLASD-----SLARLATTARAVDVPNVYVAFWRAVISL 174

```

Db 295 DFIVSNPPFKLDRSDFRDRLESDENHERFFAGIPKIKPTK--EKMEIYQLFIQHILFSL 352  
 Qy 175 KEQGRGVFIIPRSW--ANGPYRFRHLMWTAVSLDILHVPESHTKVPADTKVQENVI 232  
 Db 353 KENGKAAVLPTGFTITAKSGIDKKIREYLVENKML--AGVSMPSNIPATTGT--NWSI 407  
 Qy 233 AFSVRPOSSVVLRSVAHGEESIASVPFSAVHDEDDDKI VHPAESASVPSAARFTLA 292  
 Db 408 LFIDTKNRKVDLIDAGLGEK- IKDGNQKTVLUSCEEQKICN----- 450  
 Qy 293 DLGIGVSTGKVVDFERNROYLTDNLDSAGVPMVYQSNIRSG-----KIDWPQVW---A 342  
 Db 451 -----TTNKQAVEFVSVVGYDEIKAKNHSLSAGQVFEVKIDYVDISADBEA 498  
 Qy 343 RKQGFVA 350  
 Db 499 QKIAGFSA 506

RESULT 5  
 ID AAY88485 standard; protein; 626 AA.

XX AC AAY88485;  
 XX DT 07-AUG-2000 (first entry)  
 XX DE DraIII methylase DraIII amino acid sequence.  
 XX KW DraIII; restriction endonuclease; DraIII; two plasmid system; pHKUV5;  
 XX KW pHK77; methylase.  
 XX OS Deinococcus radiophilus.

XX FH Key Location/Qualifiers  
 XX FT Misc-difference 399 /label= Unknown  
 XX FT /note= "Encoded by NAT"  
 XX FT Misc-difference 521 /label= Unknown  
 XX FT /note= "Encoded by NTC"

XX PN US6048719-A.  
 XX PD 11-APR-2000.  
 XX PF 22-JAN-1999; 99US-00235246.  
 XX PR 22-JAN-1999; 99US-00235246.  
 XX PA (NEWE ) NEW ENGLAND BIOLABS INC.  
 XX PI Kong H, Dalton MA, Higgins LS;  
 XX DR WPI; 2000-316897/27.  
 XX DR N-PSDB; AAL15710.

XX PT Cloning and expressing an endonuclease gene by comprising two plasmid system.  
 XX PS Disclosure; Fig 4; 25pp; English.

XX CC This sequence represents the Deinococcus radiophilus (DraIII) restriction modification system methylase DraIII amino acid sequence. The DraIII endonuclease is a type II restriction endonuclease that recognises the DNA sequence 5'CACNNGTG 3', and cleaves the phosphodiester bond 5' to the first G of the recognition sequence to produce a 3 base 3' extension. The invention relates to the DraIII endonuclease gene (DraIII) and a method for its cloning and expression. The method comprises the use of a two plasmid system, using plasmids pHKUV5 and pHK77. The invention also comprises a vector containing the DNA isolated by the method, a host cell transformed by the vector, and a method for culturing the host cell and producing a DraIII restriction endonuclease

XX SQ Sequence 626 AA;  
 Query Match 6.0%; Score 147; DB 3; Length 626;  
 Best Local Similarity 19.0%; Pred. No. 1.4e-05;  
 Matches 89; Conservative 74; Mismatches 186; Indels 120; Gaps 18;  
 Qy 12 QAALGKLDPTTQAVLQGFPTPKMAATLMAFMLRVDLGRVTVLDPGAGVSLTAALVD- 70  
 Db 121 QALIG---PTIRGDKGQFFTP---KNLTDAIKILSPKPGDKIIDPACGTGFLSSCOAY 174  
 Qy 71 -RLHTEPRDVAHVAVVETDPFVVPYLRATLEECRNAYGISYDLVEGDYL--LNOGAKLD 127  
 Db 175 WELTYKOPTARYEILGIDKADNAMLSSALLEISTNGFA---KVNSDSLKFIIDNPQYE 231  
 Qy 128 GPFDLVIANPPVG-KLASDLSA-----RLATTARAVDVPNVYVAFWVRA 170  
 Db 232 EQFDIVANPPFGTKIKVDNKAAILKDYQLGHSWKVENGTLCPSRHILGAQDPQILFIELC 291  
 Qy 171 VISLKEQGRGVFIIPRSWANGPYRQFRHLMWTAVSLDILHVPESHTKVPADTKVQENVI 230  
 Db 292 VKLLKENGEMALILPEG-----VFGKSEYVWYVYLNKRG 327  
 Qy 231 IVAFSVRPOSSVVLRSVAHGEESIASVPFSAVHDEDDDKI-----V 275  
 Db 328 VPALIDCP-----RTTPOPYTDIKTNVLFKKTKEMPEKTQVAVAKRGCHDKRGRT 379  
 Qy 276 HFAESASVPSAARFTLADLIGIGVSTGKVVDFERNROYLTDNLDSAGVPMVYQSNIRSGKI 335  
 Db 380 HPSGLSVPD-----DFADIANLFEXG-----IENRIWKSVCCLKKEYRVPYRYENDDGAKL 431  
 Qy 336 DWPQVGARKQGFVAVEDVALRQLLPQGSYVVVKQTKAKEDRRRRIAADVWDSASRVALL-- 393  
 Db 432 D--NIG-----QVITIGELIRMG--VLKIRKGHE-----VGSENYGTGDIFFIR 471  
 Qy 394 DNKTYLHESQRPLEKNVARGLMNLNSTVLDQYFAPSHTQVAVAGDL 442  
 Db 472 TSDINNLFFSDP-----TNSVSEIYEMYSKKNIAAGDI 507

RESULT 6  
 ABU23716  
 ID ABU23716 standard; protein; 571 AA.  
 XX AC ABU23716;  
 XX DT 19-JUN-2003 (first entry)  
 XX DE Protein encoded by Prokaryotic essential gene #9243.  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Clostridium acetobutylicum.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US0009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JF, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX DR WPI; 2003-029926/02.  
 XX DR N-PSDB; ACA27586.



CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX  
SQ Sequence 519 AA;  
Query Match 5.5%; Score 137; DB 6; Length 519;  
Best Local Similarity 23.1%; Pred. No. 0.00011;  
Matches 74; Conservative 57; Mismatches 124; Indels 66; Gaps 17;  
QY 27 GQFTPMKAATLMASMLRVDDLRTGTVRLDPGAGVGSILTAALVDRLHTRPDVAHVAV 86  
Db 195 GEFTFQHSRLIA-QLAMHGQTSVKNLYDPAAGSGSL--LLOAKKHF-----AHII-- 244  
QY 87 ETDPRVVPYLRATLEECRN---AYGISYD---LVEGDYLLNQGAKLDGPPDLVIANPPY 139  
Db 245 -EDGFFGQEIHTTYNLAARMNFMFLHNYDKFNILMGLNLTLEPHFGDDKPFDAIVSNPPY 303  
QY 140 GK--LASDLSA-----RLATTARAVDVNPNVYAFVWRAVISLKEQGRGVFI-VPRSWANG 191  
Db 304 SVKWIGSDDEPTLINDRFPAGVLPKSKADPAFVLHLSYLSKSGRAAIVCPFGIFYRG 363  
QY 192 PYRQFRHWM-----TAVSLDLHVFSERTKVFAADTKVKQENVIYA 233  
Db 364 GAEQKIRKYLVDNNYVETVISLAPNLFPGTTIIVNLVLSKHT-----DTTQFIDASPL 419  
QY 234 FSVRQSSSVLRSVAHGEESIASVPFSAVLVHDEDDDKIVHFAESASVPSAARTLAD 293  
Db 420 F--KKEINNNIITDN--HIEQIMQ-----VFDSKDD-VHEFAKSVSFEATA---AND 463  
QY 294 LGIGYSTGKGVDFRNRQYLTD 314  
Db 464 YNLSVSS--YVEARDNREVID 482  
RESULT 9  
ABP78365  
ID ABP78365 standard; protein; 510 AA.  
XX AC ABP78365;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae amino acid sequence SEQ ID 3260.  
XX KW Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX FN WC200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX

SQ Sequence 514 AA;  
Query Match 5.5%; Score 137; DB 3; Length 514;  
Best Local Similarity 22.2%; Pred. No. 0.00011;  
Matches 73; Conservative 47; Mismatches 127; Indels 82; Gaps 14;  
QY 27 GQFTPMKAATLMASMLRVDDLRTGTVRLDPGAGVGSILTAALVDRLHTRPDVAHVAV 86  
Db 196 GEFTFQHSRLIA-RLAVHGOEKVKNLYDPAAGSGSU-----LLOAKKQFDEHI--I 245  
QY 87 ETDPRVVPYLRATLEECRN---AYGISYD---LVEGDYLLNQGAKLDGPPDLVIANPPY 140  
Db 246 BEGFFGQEIHTTYNLAARMNFMFLHNYDKFNILMGLNLTLEPHFGDDKPFDAIVSNPPY 305  
QY 141 -----KLASDLARLATTARAVDVNPNVYAFVWRAVISLKEQGRGVFI-VPRSWAN 190  
Db 306 INWIGSDDEPTLINDO--RFADPAGVLPKSKADPAFVLHLSYLSKSGRAAIVSPFGIFYR 363  
QY 191 GPYRQFRHWM-----TAVSLDLHVFSERTKV-FADTKVKQEN 229  
Db 364 GAEQKIRKYLVEGNYVETVIALAPNLFYGTGIAVNLVLSKHKONTDIQFIDAS----- 418  
QY 230 VIVAFSVRQSSSVLRSVAHGEESIASVPFSAVLVHDEDDDKIVHFAESASVP-----S 285  
Db 419 ---GFFKXETNNVLI-----EEHIAEIVKL-----PADKADVPHIAQN 454  
QY 286 AARFTLADLGIGYSTGKGVDFRNRQYLTD 314  
Db 455 AAQQTVDKNGYNLAVSSVVEAEDTREIID 483  
RESULT 8  
ABM70049  
ID ABM70049 standard; protein; 519 AA.  
XX AC ABM70049;  
XX DT 20-NOV-2003 (first entry)  
XX DE Photorhabdus luminescens protein sequence #3146.  
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
XX detection; food; gene expression; plant; animal; microorganism; toxin;  
XX antibiotic; biopesticide; virulence factor; disease model; plague;  
XX whooping cough.  
XX OS Photorhabdus luminescens.  
XX FN WC200294867-A2.  
XX PD 28-NOV-2002.  
XX PF 07-FEB-2002; 2002WO-IB003040.  
XX PR 07-FEB-2001; 2001FR-00001659.  
XX PA (INSP ) INST PASTEUR.  
XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
XX PI Buchrieser C;  
XX WI PI; 2003-148459/14.  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX Claim 2; SEQ ID NO 3146; 1205pp; French.  
XX The invention relates to the isolation of genes and their encoded  
XX proteins from Photorhabdus luminescens. The isolated sequences are  
XX sources of probes and primers for detecting the genome of P. luminescens  
XX and related species; to study polymorphisms; for gene analysis and for

PI Fontana MR, Pizza M, Masignani V, Monaci E;  
XX  
DR WPI; 2003-058415/05.  
XX N-PADB; ABZ93335.

New protein from Neisseria gonorrhoeae, useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection.

Discloure; Page 427; 815pp; English.

The present invention relates to proteins from Neisseria gonorrhoeae. Also disclosed are the nucleic acid molecules encoding the proteins and antibodies that specifically bind to the proteins. The composition comprising the protein, nucleic acid or antibody is useful for the manufacture of a medicament for treating or preventing N. gonorrhoeae infection, this may be in the form of a vaccine or gene therapy. Sequences given in records ABP76736-ABP81046 represent nucleic acid molecules of the invention

XX Sequence 510 AA;

Query Match                5.5%; Score 135; DB 6; Length 510;  
Best Local Similarity     22.0%, Pred. No. 0.00017;  
Matches      76; Conservative    52; Mismatches 128; Indels    90; Gaps    15;

Qy    27   QGFPTPKAAATLMASMLRVDLRTGTVRLDPGAGVGSLLTAALVLRHLPEDVAHVAV    86  
Db    :|||: |:||: ||:: |||: |||:                :    :    :  
192   GEFTTPQSVMKIIA-RLAVHQEKVNKIYDPAKGSL-----LIQAKKQFDEHI--I    241

Qy    87   ETDFPVVPYLRAITVEECRN---AYGISYD---LVESGYLLNQGAKLDGPFDLVIANPPYG    140  
Db    :|||: |:||: ||:: |||: |||:                :    :    :  
242   EGFPFGOEINHTTTLARMMFLHNKNKFHIELGDTLTNPKLKDSKPDAVVSNPPYS    301

Qy    141   -----KLADPSLARLATARAADVNVVAFWRAVISLKBOGRGVFI-VPRSWAN    190  
Db    :|||: |:||: ||:: |||: |||:                :    :    :  
302   IDWTSGDDPTLINDD--RFAPAGVLAPKSKADFALIHNLNYLSGRGRAAIIVSPFGIFYR    359

Qy    191   GPYYQRFRHWIM-----TAVSLDLTHVPESRTKVFAETDKVKQENVIV    232  
Db    :|||: |:||: ||:: |||: |||:                :    :    :  
360   GGAEQIKROYLVENGYVETVIALPNLFYGTCTIAVNLILVLSKHK-----DNTDI    408

Qy    233   AF-----SVRPQSSSVLSRSVAHGEEISASSYPFSALVHDEDDDKIVHFAESASP---    284  
Db    :|||: |:||: ||:: |||: |||:                :    :    :  
409   QFIDASGFFFKKETNNVT-----EBHIAEIVKL-----PADKADVPHTA    448

Qy    285   -SAREFTLADLGIGVGSTCKVDFNRQYLTD----NLDASGVVPMV    325  
Db    :|||: |:||: ||:: |||: |||:                :    :    :  
449   ONAAQQTKDNGYNLAVSRYVEAEDTREVIDIRQLNAEISETVAKI    494

RESULT 10  
AAV75217

ID AAV75217 standard; protein; 514 AA.

XX AC AAV75217;  
XX DT 12-SEP-2003 (revised)  
XX DD 21-MAR-2000 (first entry)

Neisseria gonorrhoeae ORF 605 protein sequence SEQ ID NO:1908.

XX DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
KW KW antibacterial; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WO9957280-A2.  
XX PD 11-NOV-1999.  
XX PF 30-APR-1999; 99WO-US0009346.  
XX XX

XX AC ABU37303;  
XX DT 23-OCT-2003 (revised)  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #22830.  
XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX KW Neisseria gonorrhoeae.  
XX OS  
XX PN W0200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-029926/02.  
XX DR N-PSDB; ACA41173.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 65227; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
XX CC standardise OS field)  
XX CC Sequence 514 AA;

Query Match 5.5%; Score 135; DB 6; Length 514;  
Best Local Similarity 22.0%; Pred. No. 0.00016;  
Matches 76; Conservative 52; Mismatches 128; Indels 90; Gaps 15;  
QY 27 GQFTPMKAATLMASMLRVDDLRGTVRVLPDGAGVGSLLAALVDRHLHTEPDVAHVAV 86  
DB 196 GEFPTQSVSKLIA-RLAVHGOEKVKNKIYDPACGSSL-----LLQAKKQFDEHI--I 245  
QY 87 ETDFVFPVYLRLTECRN---AYGISYD---LVGEDYLLNQAKLOGPFDLVIANPPYG 140  
DB 246 EEGFFGOEINHHTTYNLARMNMFHNVNYPKHIELGDTLTNPCLKSKDFDAVVSPPYS 305  
QY 141 -----KLASDLARLATTARADVDPNVYVAVFVRAVISLKEQSGVFI-VPRSWAN 190  
DB 306 IDWIGSDPTLTINDD--RFAPAGVLAPKSKADPAFILHALNLYLSGRRAAIVSPGFIYR 363  
QY 191 GPYYRQFRHMLM-----TAVSLDILHVFESRTKVPADTKVQENVIV 232  
DB 364 GGAEQKIRQYLVVEGNYVETVIALAPNLFYGTCTIAVNILVLSKHK-----DNTDI 412  
QY 233 AF-----SVRPQSSSVLSRSVAHGEESIASVPSFSLVHDEDDDKIVHFAESASVP--- 284  
DB 413 QFIDASGFFKKTNNVLT-----EEHIAEIVKL-----FADKADVPPIA 452  
QY 285 -SAARFTLADLIGVSTGKVDFNRQYLTLD-----NLDASGVVPMV 325  
DB 453 QNAAQOTVKDNGYNLAVSSYVEAEDTREVIDIRQLNAEISETVAKI 498  
RESULT 12  
AAB07560  
ID AAB07560 standard; protein; 2162 AA.  
XX AC AAB07560;  
XX DT 20-OCT-2000 (first entry)  
XX DE Protein encoded by the bleomycin (BLM) gene cluster ORF30.  
XX KW BLM gene cluster; bleomycin gene cluster; polyketide metabolite;  
XX KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;  
XX KW thiazoline; bithiazoline; microbial metabolite; sugar.  
XX OS Streptomyces verticillus.  
XX PN W0200040704-A1.  
XX PD 13-JUL-2000.  
XX PF 06-JAN-2000; 2000WO-US0000445.  
XX PR 06-JAN-1999; 99US-0115435P.  
XX PR 05-FEB-1999; 99US-0118848P.  
XX PR 05-JAN-2000; 2000US-00477962.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI Shen B, Du L, Sanchez C, Chen M, Edwards DJ;  
XX PI WPI; 2000-465974/40.  
XX DR N-PSDB; AAA58471.  
XX CC New bleomycin gene cluster components useful for peptide and/or  
XX CC polyketide metabolites, especially bleomycin, production and for  
XX CC chemically modifying biological molecules.  
XX PS Disclosure; Page 97-137; 162pp; English.  
XX CC AAB07556-78 represent proteins encoded by open reading frames (ORFs) 8 to  
XX CC 30 of the BLM (Bleomycin) gene cluster. The proteins encoded by the gene  
XX CC cluster are useful for producing peptides and/or polyketide metabolites,  
XX CC especially bleomycin or bleomycin analogues. They are also useful for  
XX CC chemically modifying biological molecules to produce branched methyl



CC groups, and for coupling amino acids and fatty acids. They may be reacted  
CC with an apo-carrier protein and coenzyme A to produce a holo-carrier  
CC protein. The BLM gene cluster or catalytic domains can be used  
CC individually or collectively to produce thiazolidine, thiazoline,  
CC bithiazoline and bithiazoline-containing microbial metabolites. The BLM  
CC gene cluster may also be used to produce sugars  
xx  
SQ Sequence 2162 AA:

Query Match	5.5%;	Score 135;	DB 3;	Length 2162;
Best Local Similarity	22.6%;	Pred. No. 0.002;		
Matches 115;	Conservative 47;	Mismatches 189;	Indels 158;	Gaps 23;
QY	4	IASNETRQALGKLDITTOQVLQCOFTPMKAATLMSMLRVDDLRGTVRVLDPGAGVGS	63	
Db	708	IAAUSRRDRLTKLTPTHLDDVNNQLITP-----DELRGAVRTLTVVGGEA--	752	
QY	64	LTAALVDRLHTERPDAVHVAVVAVETDPFVFPYLKATLEECRNAGISVDLVEG--DYLLNQ	122	
Db	753	-----VRAESLEFP-----RAGTRVNVNYPGETVVGSVVAHVDDA	788	
QY	123	GAKLDGGFDLVIANPPYVKGLASDSLARLATTARAVDVPNVYVAFWRAVVISLKEQGRGVF	182	
Db	789	ATPRTGPV-----PIGRPIANTVHLLDQRRFPDVGVGELMI-----GGAGVA	833	
QY	183	--TVPSWANGPY-----RQRRHMLTAVSLDILHVPESRTKVFADTKVKQENV	230	
Db	834	DVGLGPELTGTERFLPSDYPDPDGGVRYRTGDLARRRADGTLEYLGR-----DAQVKIRGV	889	
QY	231	IVAFSVRPOSSVVLSRSVAHGSESIASSVPFSALVHDEDDKIVHPAESASVPSAARFT	290	
Db	890	----RVPEAPTEAVLASHPGVGQAVVARI-----DEDGPR-----SSPLAGELT	930	
QY	291	LADLGIGVSTGKVVDFFNRQ-----YLTNLDASGV-----VPMVYQSNIRS	332	
Db	931	L-----TGYVVVPARGAQAAPHEELIAYCERELPEHFVPVAVLVTLDALPVTGCHKIDR	982	
QY	333	GKIDWPQVGARKQGVAVVEDVALROLLPGQSVVVKRQTAKEDERRRVIAAVWDGASRVA	392	
Db	983	GALPKPHARARGAAVYA-----PRTATEIILAAATVAKVLGVVERVG	1023	
QY	393	LDNKNTYL---HESORPL---EKNVARGMLMLNSTVLDOY-----FRPSGHTQVNVAGDL	442	
Db	1024	IDD--NYFVLGGDSIRSVMVSAQAQARGVEV-----TVADLHRHTPVRAAAHLDAARE--DL	1076	
QY	443	RLRP-----FLCREDLIILAKKVVDPGLP	465	
Db	1077	PRTPVTEPPFALISAEDRALVPDDVEDAFPP	1105	

RESULT 13	
AAAY75219	
IID	AAAY75219 standard; protein; 514 AA.
XX	
AC	AAAY75219;
XX	
DT	21-MAR-2000 (first entry)
XX	
DE	Neisseria meningitidis ORF 605 protein sequence SEQ ID NO:1912.
XX	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW	antibacterial; gene therapy.
XX	
OS	Neisseria meningitidis.
XX	
PN	WO9957280-A2.
XX	
IPD	11-NOV-1999.
XX	
PPF	30-APR-1999; 99WO-US009345.
XX	
PR	01-MAY-1998; 98US-0083758P.

PR	31-JUL-1998;	98US-0094869P.	
PR	02-SEP-1998;	98US-0098994P.	
PR	02-SEP-1998;	98US-0099062P.	
PR	09-OCT-1998;	98US-0103749P.	
PR	09-OCT-1998;	98US-0103794P.	
PR	09-OCT-1998;	98US-0103796P.	
XX	25-FEB-1999;	99US-0121528P.	
PA	(CHIR ) CHIRON CORP.		
PA	(GENO-) INST GENOMIC RES.		
XX			
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;		
PI	Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;		
PI	Tettelin H, Venter JC;		
XX			
DR	WPI; 2000-062150/05.		
DR	N-PSDB; AAZ53981.		
XX			
PT	Novel Neisserial polypeptides predicted to be useful antigens for		
PT	vaccines and diagnostics.		
XX			
PS	Claim 2; Page 963; 1453pp; English.		
XX			
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941		
CC	represent novel <i>Neisseria meningitis</i> and <i>N. gonorrhoeae</i> polynucleotides		
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent		
CC	PCR primers used in the exemplification of the present invention. The		
CC	polypeptides, the polynucleotides, antibodies and compositions of the		
CC	invention can be used as vaccines, as diagnostic reagents, and as		
CC	immunogenic compositions. The polypeptides can be used in the manufacture		
CC	of medicaments for treating or preventing infection due to <i>Neisserial</i>		
CC	bacteria (e.g. <i>meningitis</i> and <i>septicaemia</i> ), to detect the presence of		
CC	<i>Neisseria</i> bacteria, or to raise antibodies. They may also be used to		
CC	screen for agonists or antagonists, which may themselves have use as		
CC	antibacterial agents. The polynucleotides of the invention may also be		
CC	used in gene therapy protocols		
XX			
XX	Sequence 514 AA;		

[illegible]





Search completed: October 1, 2004, 16:12:55  
Job time : 67.6786 secs

P Lank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:09:01 ; Search time 19.8245 Seconds  
(without alignments)  
1265.619 Million cell updates/sec

Title: US-10-668-047-2  
Perfect score: 2469  
Sequence: 1 VHPFATETTRQAALGKLDP.....QETLDVAVARLFCIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PGTUS COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	136.5	5.5	1309	4	US-09-975-413A-10	Sequence 10, Appl
2	122	4.9	534	4	US-09-107-532A-6592	Sequence 6592, Ap
3	120.5	4.9	535	4	US-09-107-532A-6593	Sequence 6593, Ap
4	117.5	4.8	507	4	US-09-252-991A-26552	Sequence 26552, A
5	116.5	4.7	414	4	US-09-252-991A-21182	Sequence 21182, A
6	114.5	4.6	597	4	US-09-252-991A-31010	Sequence 31010, A
7	113.5	4.6	1009	4	US-09-693-146-4	Sequence 4, Appl
8	111.5	4.5	549	4	US-09-693-146-2	Sequence 2, Appl
9	109.5	4.4	672	4	US-09-252-991A-32554	Sequence 32554, A
10	109	4.4	1007	4	US-09-957-005-9	Sequence 9, Appl
11	106.5	4.3	504	4	US-09-134-001C-2980	Sequence 2980, Ap
12	106	4.3	406	4	US-09-071-035-114	Sequence 114, App
13	105	4.3	557	4	US-09-134-000C-4354	Sequence 4354, Ap
14	104.5	4.2	515	4	US-09-489-039A-10635	Sequence 10635, A
15	101	4.1	202	4	US-08-311-731A-13	Sequence 13, Appl
16	100.5	4.1	296	4	US-09-134-001C-3320	Sequence 3320, Ap
17	99.5	4.0	338	4	US-09-489-039A-11940	Sequence 11940, A
18	99.5	4.0	604	2	US-08-746-283-32	Sequence 32, Appl
19	99.5	4.0	604	2	US-08-746-257A-30	Sequence 30, Appl
20	99.5	4.0	605	1	US-08-333-802-2	Sequence 2, Appl
21	99.5	4.0	605	4	US-09-347-878-46	Sequence 46, Appl
22	99	4.0	528	4	US-09-071-035-44	Sequence 44, Appl
23	99	4.0	547	4	US-09-071-035-42	Sequence 42, Appl
24	98.5	4.0	513	4	US-09-489-039A-14224	Sequence 14224, A
25	98.5	4.0	554	4	US-09-540-236-3778	Sequence 3778, Ap
26	98.5	4.0	618	3	US-09-299-378-4	Sequence 4, Appl
27	97.5	3.9	301	4	US-09-489-039A-10251	Sequence 10251, A

28 96.5 3.9 926 4 US-09-252-991A-32551 Sequence 32551, A

29 95 3.8 800 4 US-09-328-352-5709 Sequence 5709, Ap

30 95 3.8 3241 4 US-03-841-786-1 Sequence 1, Appl

31 95 3.8 3892 4 US-09-328-352-5503 Sequence 5503, Ap

32 94.5 3.8 287 4 US-09-328-352-7035 Sequence 7035, Ap

33 94.5 3.8 336 3 US-09-105-537-27 Sequence 27, Appl

34 94.5 3.8 351 4 US-09-134-000C-4436 Sequence 4436, Ap

35 94.5 3.8 11877 3 US-09-105-537-6 Sequence 6, Appl

36 93.5 3.8 398 4 US-09-134-001C-5523 Sequence 5523, Ap

37 93.5 3.8 927 4 US-09-841-786-3 Sequence 3, Appl

38 93.5 3.8 3074 4 US-09-543-681A-5508 Sequence 5508, Ap

39 93 3.8 253 4 US-09-180-109A-35 Sequence 35, Appl

40 93 3.8 254 4 US-09-180-109A-31 Sequence 31, Appl

41 93 3.8 664 4 US-09-328-352-6193 Sequence 6193, Ap

42 93 3.8 828 4 US-09-198-452A-323 Sequence 323, App

43 92.5 3.7 197 4 US-09-252-991A-27889 Sequence 27889, A

44 92.5 3.7 414 4 US-09-252-991A-27975 Sequence 27975, A

45 92.5 3.7 419 3 US-08-258-287B-57 Sequence 57, Appl

## ALIGNMENTS

## RESULT 1

US-09-975-413A-10  
; Sequence 10, Application US/09975413A  
; Patent No. 6593122

; GENERAL INFORMATION:  
; APPLICANT: Maunus, Robert

; APPLICANT: Xu, Shuang-Yong

; APPLICANT: Benner, Jack

; TITLE OF INVENTION: Method For Cloning And Expression of BserI Restriction Endonuclease

; TITLE OF INVENTION: BserI Methylase In E. coli

; FILE REFERENCE: NEB-193

; CURRENT APPLICATION NUMBER: US/09/975,413A

; CURRENT FILING DATE: 2001-10-11

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Patent version 3.1

; SEQ ID NO 10

; LENGTH: 1309

; TYPE: PRT

; ORGANISM: Bacillus species R (CAMB2669)

US-09-975-413A-10

Query Match 5.5%; Score 136.5; DB 4; Length 1309;

Best Local Similarity 24.4%; Pred. No. 9.9e-05;

Matches 63; Conservative 40; Mismatches 94; Indels 61; Gaps 13;

QY 17 KLDPTTQAVLIGQFPTPKAATLWASMLRVDLRLGTV--RVLDPGAGVGSILTAALVDRLHT 74

Db 630 QINPERQKRLQVFTSGPLAELATF-----AEGSTASSVIDMCGGDMLTAVNSINS 683

QY 75 ERPDVAHVHVAVETDPVPVYLRLATLEECRNAYG----ISYDLVEGD-YLINQAKLD-GP 129

Db 684 K-----ANLSGIDIDPI-----AMNKCIDRLGNKKSLLDILIGSAFQSWNTIKQLKLS 731

QY 130 FDLVIANPY-----GKLSDSLRLATLTAR-----AVD-----158

Db 732 FDLVITNPVRYQSLSSKLEGLVDLIPDSETVNDLLEVVSQLDHLHRDKVEFTVKS 791

QY 159 ---VPNVYAVFVRAVISLKEQGRGVFIIVPRSWANGPYRQFEHMLTAVSLDLHVFFS 215

Db 792 YGSLSDLAIVPSILCMLTSVGHLMAMVVPESWLNRYAHPHYLLLLFKIKWVVDYN 851

QY 216 RTKVFADTKVKQENVIVA 233

Db 852 RT-WFKDAQVK-TNLVVA 867

## RESULT 2

US-09-107-532A-6592

; Sequence 6592, Application US/09107532A

; Patent No. 6583275

;; GENERAL INFORMATION:  
;; APPLICANT: Lynn A Doucette-Stamm and David Bush  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
;; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
;; NUMBER OF SEQUENCES: 7310  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
;; STREET: 100 Beaver Street  
;; CITY: Waltham  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02354  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: CD-ROM ISO9660  
;; COMPUTER: PC  
;; OPERATING SYSTEM: <Unknown>  
;; SOFTWARE: ASCII  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/107,532A  
;; FILING DATE: 30-Jun-1998  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: 60/085,598  
;; FILING DATE: 14 May 1998  
;; APPLICATION NUMBER: 60/051571  
;; FILING DATE: July 2, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Ariniello, Pamela Deneka  
;; REGISTRATION NUMBER: 40,489  
;; REFERENCE/DOCKET NUMBER: GTC-012  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (781)893-5007  
;; TELEFAX: (781)893-8277  
;; INFORMATION FOR SEQ ID NO: 6592:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 534 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: YES  
;; ORIGINAL SOURCE:  
;; ORGANISM: Enterococcus faecium  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (B) LOCATION 1...534  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6592:  
US-09-107-532A-6592

Query Match 4.9%; Score 122; DB 4; Length 534;  
Best Local Similarity 24.1%; Pred. No. 0.00075;  
Matches 70; Conservative 35; Mismatches 116; Indels 70; Gaps 15;  
QY 27 GQFFTPMKAATLWASMLRVDDL-RGTVRLDPGAGVSLTAALVDELHTRPD-VAVHV 84  
Db 198 GEFYTPHVMDSMAQIVTLTDKERRFFSVFDTMTSGSL--MLNVRNYLTHPDNRYKHQ 255  
QY 85 AVETDPFVVPYLRATLEECRNAYGI---SYDLVEGDYLLNOGAKLDGP--FDLVIANPP- 138  
Db 256 ELNNTTYNLAKNMLIL-----HGVDAAEMNLRNGD-TLNKQWPTDEPTTFDAVVMNPPY 308  
QY 139 -----YGLKASDSLARLATTARAVDVPNVYVAFWVRVAVISLKEQGRG 180  
Db 309 SANWSADTTFLDSDRENRYGKLAPSKADF-----AFLHGFVHLKETGT 354  
QY 181 VFIWPSWANGPYRQFRHMLMTAVSLD---ILHVFESRTKVPADTKVQENIVAFSVR 237  
Db 355 AIVLP-----HGVLFGRGAEGVIRQKLEDGSIYAVIGMPANLPFGTSI--PTTVIVLKN 408  
QY 238 PQSSSVVL---SRVAHG-----EESIASVFPFSAVHDEDDDKIVHFA 278  
Db 409 ROTRDVLFDASREFVKGKQNKLSNIQKIL--ETVAKDKVKAHLA 457

RESULT 3

US-09-107-532A-6593  
;; Sequence 6593, Application US/09107532A  
;; Patent No. 6583275  
;; GENERAL INFORMATION:  
;; APPLICANT: Lynn A Doucette-Stamm and David Bush  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
;; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
;; NUMBER OF SEQUENCES: 7310  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
;; STREET: 100 Beaver Street  
;; CITY: Waltham  
;; STATE: Massachusetts  
;; COUNTRY: USA  
;; ZIP: 02354  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: CD-ROM ISO9660  
;; COMPUTER: PC  
;; OPERATING SYSTEM: <Unknown>  
;; SOFTWARE: ASCII  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/107,532A  
;; FILING DATE: 30-Jun-1998  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: 60/085,598  
;; FILING DATE: 14 May 1998  
;; APPLICATION NUMBER: 60/051571  
;; FILING DATE: July 2, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Ariniello, Pamela Deneka  
;; REGISTRATION NUMBER: 40,489  
;; REFERENCE/DOCKET NUMBER: GTC-012  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (781)893-5007  
;; TELEFAX: (781)893-8277  
;; INFORMATION FOR SEQ ID NO: 6593:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 535 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHETICAL: YES  
;; ORIGINAL SOURCE:  
;; ORGANISM: Enterococcus faecium  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (B) LOCATION 1...535  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 6593:  
US-09-107-532A-6593

Query Match 4.9%; Score 120.5; DB 4; Length 535;  
Best Local Similarity 23.0%; Pred. No. 0.0011;  
Matches 79; Conservative 39; Mismatches 135; Indels 91; Gaps 18;  
QY 27 GQFFTPMKAATLWASMLRVDDL-RGTVRLDPGAGVSLTAALVDELHTRPD-VAVHV 84  
Db 198 GEFYTPHVMDSMAQIVTLTDKERRFFSVFDTMTSGSL--MLNVRNYLTHPDNRYKHQ 255  
QY 85 AVETDPFVVPYLRATLEECRNAYGI---SYDLVEGDYLLNOGAKLDGP--FDLVIANPP- 138  
Db 256 ELNNTTYNLAKNMLIL-----HGVDAAEMNLRNGD-TLNKQWPTDEPTTFDAVVMNPPY 308  
QY 139 -----YGLKASDSLARLATTARAVDVPNVYVAFWVRVAVISLKEQGRG 180  
Db 309 SANWSADTTFLDSDRENRYGKLAPSKADF-----AFLHGFVHLKETGT 354  
QY 181 VFIWPSWANGPYRQFRHMLMTAVSLD---QFRHMLMTAVSLDILHVFESRTKVPADTKVQENIVAFS 235  
Db 355 AIVLP-----HGVLFGRGAEGVIRQKLEDGSIYAVIGMPANLPFGTSI--PTTVIVL 406  
QY 236 VFPQSSSVVL---SRVAHGEEISASSVFPFSAVHDEDDDKIVH-FAESASVPSAARFTL 291  
Db 407 KNRQTRDVLFDASREFVKGKQNKLS-----EENIQKILETYAERKDKVKAHLAT 458

QY 292 ADLIGVSTGKVDPRNQY-----LTDNLDSAGVPMVYQSN 329  
Db 459 FD-----BIKENDYLNIPRYVDTFEEBEPIDMVHGN 491

## RESULT 4

US-09-252-991A-26552  
; Sequence 26552, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26552

; LENGTH: 507  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26552

Query Match 4.8%; Score 117.5; DB 4; Length 507;  
Best Local Similarity 24.3%; Pred. No. 0.002;  
Matches 107; Conservative 52; Mismatches 196; Indels 95; Gaps 23;

QY 50 GTVRVLDPCAGVGLTAALVDRLHTERPDVAVHVAVETDPFVVPVYLRATLEECRNAYGI 109  
Db 64 GRMKICPCAGEG--VALAAEAHILGRD-KVQALAVEYDRADHARGLLERVLHS--- 116

QY 110 SYDLVEGYLLNOGAKLDGDFDLVIANPPYKGLASDLARLATTARADVDPNVYVAFWVR 169  
Db 117 --DLF--DTMISRQS-----FGLLWLNPPYGLDLVADHSG--ASQYQSGRRRLEKAFYQR 165

QY 170 AVISLKEQGGGVFIVPRSWANGPY--RQFRHWTAVSLDLHVFEGRTKVFATKVK 226  
Db 166 CUPLLQYGMVLLIV-----PHYVLDLDTLGTWLSN-----HFTGLRMVAAADPTFK 211

QY 227 QENVIVAFSVRPO-----SSSWLSRSVAHG-----EESIASVSPFSALVHDEDDKI 274  
Db 212 Q--VVIFGRVQRDLARADANQVSRQAIGAGQEKAEETPAWPWPYVVLPAATSEL 268

QY 275 VIFAESASVPSA-----ARFTLADLIGVSTGKVDPRNQYLTNDLDSAGVV 322  
Db 269 EHFYRVTLPEQFAGEIQRGLWDPDFNLHFAQAGLQPRPPVRELRSWHLALAL-AAGAI 327

QY 323 PMVYQSNIRSGKI-----DWPQVGARKPGGFVAVEDVALRQLLPQGSYVVVKRQTAKER 377  
Db 328 SGVRS--KSGRLLVVGDTYKDKARKTE-FTEDED-----GNITEVRLIT---DR 372

QY 378 RRVIAAVND-----GASRV-ALDKNKTYLHFSQRPLEKNVARGMLNLSVLDQYFA 430  
Db 373 FIPIIRAWEMTPSSVQNGRVLISSAAATEAEPEQPEAPEPLIFSPGQVVM-----TA 428

QY 431 FSGHTQVNGADLRRUPLFLCR 450  
Db 429 AVSHL-VETGQLNPAPLLNR 447

## RESULT 5

US-09-252-991A-21182  
; Sequence 21182, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21182

; LENGTH: 414  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21182

Query Match 4.7%; Score 116.5; DB 4; Length 414;  
Best Local Similarity 24.4%; Pred. No. 0.0019;  
Matches 77; Conservative 31; Mismatches 106; Indels 101; Gaps 16;

QY 17 KLDPTTQAVLGQF--FTPMKAATILMASM-----LRVDD 47  
Db 22 RLDPPSEARLGVLPSPFIPANQETSMALMFPRLARNFAKNGYYPTDEPTLRALNALMPSD 81

QY 48 LRGTVRVLDPCAGVGLTAALVDRLHTERPDVAVHVAVETDPFVVPVYLRATLEECRNAY 107  
Db 82 --GPMCLIDPCAGEGVALAEASHALGSEQ-----AKAFAVEFDAERARHARGLDHCUHA- 134

QY 108 GISYDLVEGYLLNOGAKLDGDFDLVIANPPYKGLASD-----SLARLATTARAV 157  
Db 135 ---DLM--DTMISKQS-----FGLLWLNPPYGLSKDVNGNIGYQCGGRARLEK----- 178

QY 158 DPNVYVAFWVRVAVISLKEQGGGVFIVPR-----SWANGPYVYQFRHWTAVSLDI 209  
Db 179 -----LFYQSLSLQYSGVLVFPVGVLDLAEVLGMLT-----RHY-----TDLRI 220

QY 210 LHVFEGRK--VFADTKVKQENVIVAFSVRPSQSSVV--LSRSVAHGE---ESIASVVPF 262  
Db 221 YRAVETQFKQVIFGRVRQR-----ELAPGVKAVRNLLLOVGLGEVAEELPSEWPF 274

QY 263 SALVHDEDDDKIVHF 277  
Db 275 LPYIVPASPAEPEHF 289

## RESULT 6

US-09-252-991A-31010

; Sequence 31010, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31010  
; LENGTH: 597

; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31010

Query Match 4.6%; Score 114.5; DB 4; Length 597;  
Best Local Similarity 23.4%; Pred. No. 0.0056;  
Matches 44; Conservative 30; Mismatches 75; Indels 39; Gaps 9;

QY 27 GQFTPMKAATILMASMLRVD--DLRGTVRVLDPCAGVGLSTAAALVDRLHTE-----R 76  
Db 175 GEFTFTRAITAFMAD--RVNPRLDKRET--VMDPACGTGGFLTAIDHFRNLQSLKSSAE 230



QY 198 -----RHLMATAVS-----LDI 209  
Db 217 LGELTHYIGIEDITONKLAQOHEKLAIRDNLGLANRHYFIGALEERLESSGDRPLSL 276  
QY 210 LHVFSRTRKVPADTKVKQ--ENVIVAFVRPOSS-----SVVLSRSVAHG 252  
Db 277 LLVDIDNFRINDSLGHOTGDKLLVSLARRLRSLCLGDGATLARFASNEFAVLLDDTAVEK 336  
QY 253 EESIASVPFSSALVHDEP-----DDKIVHFAESASVPSAARF-----TLADLGIGVSTG 301  
Db 337 GESIAAQV-----LHMLDKPLFVDNQLINITGSLASAPQHGCDPQTLKMYAGLALHKA 391  
QY 302 KVDPRNRQYITDNLDSAGVPMVYQSN-----IRSGK-----I 335  
Db 392 KANGKHQVQVTEALTAEASVKLFVESNLRRALAQNELAVHYQPKCLRSQQLGLEALL 451  
QY 336 DW--POVGARKEQGFVAV-EDVALRQLLPQGSYVVVKQTAKEDRRRRVIAAVWDGASRVA 392  
Db 452 RWQHPKGMIRPDRFISVAETGL--IVPIGKWI--REACROARELAELAGL--GELQIA 505  
QY 393 LD-----NKTNYLHESORP---LEKNVARGMLMLNLSVLDQYFRAFS-----432  
Db 506 INLSPKQFTDPDLVGSIAALIHENIPASQLELELTSLDADTDTROQLERKLSGLT 565  
QY 433 -----GHTQVWAGDLRLPFLCREDLILL-----AKVVPDGLPQDQETLDVAV-----475  
Db 566 LAMDDFGTGYSSLSY--LKKFPI---DVIKIDRSFIKIDPSODDMEITSAVIAMAHNL 619  
QY 476 --RLFCFEPESA 485  
Db 620 KLVVAEGVESA 631

## RESULT 10

US-09-957-005-9  
; Sequence 9, Application US/09957005  
; Patent No. 6596524  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Zhenyu  
; APPLICANT: Zhou, Jing  
; APPLICANT: Xu, Shuang-yong  
; TITLE OF INVENTION: Method For Cloning And Expression of BsmAI Restriction  
; FILE REFERENCE: NEB-192  
; CURRENT APPLICATION NUMBER: US/09/957,005  
; PRIOR FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1007  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus A664  
US-09-957-005-9

Query Match 4.4%; Score 109; DB 4; Length 1007;  
Best Local Similarity 18.6%; Pred. No. 0.051;  
Matches 105; Conservative 102; Mismatches 215; Indels 142; Gaps 28;  
QY 17 KLDPTTQAVLGQFFTPMKAATLM-----ASMLRVDDLRTGTVRLDPGAGVSLTAALVD- 70  
Db 36 QMDPLERKETSFTALELTDVMMQBLVSYLKSXDKITELKFLPCVGTGNGVFSYLKE 95  
QY 71 ----RLHTERPDAVHVAVE-----TOPFVVPYLRATL-EECRNAYGISVD 112  
Db 96 ISKLQLHKEQIETLNNIYVADINQTAALLEYKLLSKFAKLYPDIDLSEYFNHSGSAL 155  
QY 113 LVE-----GDYLLNQAKLDGPFDLVIANPPY-----GKLASD-----145  
Db 156 LIDVAEQPEYIKITDVFDEVKEG-----FDIVTNPYPNKAEGQVSNLVEYI 209  
QY 146 SLARLATTAARV-----DVPNVYVAFWVRVAVISLKEQGRGV--FIVPRSWANGPPY 194

Db 210 DRARYAEIKMKVRIFNYSTDGLNLYKLF-VEEIIDKYANPNGFVSLILPSSILTDKTC 268  
QY 195 RQFR-HWMTAVSLDILHVFESRTKVFADTKVKQENVIVAFSVRPQSSSVLSRSVAHGE 253  
Db 269 TKLRTHMLVDSNILLSIKMINEGSGVIDA---QQALSAILIQKGRTESIKVTKDYSNNP 324  
QY 254 ESIASSVPFSAVHDEDDDKI--VHFABASVPSAARF-TLADLGIGVSTGKVDFRNRQ 310  
Db 325 NQI--TDINMEDILNENTGNALFAINNHEYFYLKQLRKFVVPVKDLDF-----IINLRGEL 377  
QY 311 YLTNLDASGVPMVYQSNIRSGKIDWPQVGARKPQGFVAVEDVALROLL-----PQ 362  
Db 378 DLTANKDS--IV-----NIDTG--YPLLRGRN-IGYIEILDTCSGFVSKDFIENSKK 425  
QY 363 GSYVVVVR-----QTKEDRRRRVIAAVWDGASRVALDNKTNLYLHESORPLEKRVARGLM 416  
Db 426 SRYIKERKIVCQVVNNMKERRVTFALV--EENVVLGNSCNFI--SVMDNDYNDILYAI 480  
QY 417 LWL-NSTVLDQYFRAFSGHTQVWAGDLRLRP-----FLCREDLILIAK 458  
Db 481 LGLENTSIIIMWLFKLTSSNNHNNVEIDCFVPPIGSPYLNKISNLVKKYLKNSKOSLLEK 540  
QY 459 V-----VPDGLPQDQETLDVAV 473  
Db 541 IEEVAYIAYGIREAKEDNEKD KDI 564

## RESULT 11

US-09-134-001C-2980  
; Sequence 2980, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2980  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2980

Query Match 4.3%; Score 106.5; DB 4; Length 504;  
Best Local Similarity 27.1%; Pred. No. 0.029;  
Matches 57; Conservative 30; Mismatches 80; Indels 43; Gaps 16;  
QY 4 IASTETRRQAAAG-----KLDPTTQAVLGQFFTPMKAATLMASMLRVDDLRTGTVRL 55  
Db 137 VDGTESRKQDVLGRVVEYFIAPASAEGRNAGEFTYFPSSIVKLLVEM--IEPYKG--RIY 192  
QY 56 DFCAGVGSLL---TAALVDRHLHTERPDVAVHVAVEDTPFVVPYLRATLEECRNAYGISVD 112  
Db 193 DPCGSGGMFVQSERFVEKHQGLDDIAY--GOESNPTTWKLAQNLA-----IRGIDND 246  
QY 113 LVE--GDYL---INQAKLDGPFDLVIANPPYKGLASD-SLARLATTAAR-AVDVP---NV 162  
Db 247 LGERNSDTFPHDLHKGLKA---DYILANPPFN--ASDWGQERLLDDYRWQFGIPKPGA 300  
QY 163 YVAFWVRVAVIS-LKEQGRGVFIVPRSWANG 191  
Db 301 NYA-WIEHMISKLAENGTAGFVL---ANG 325

## RESULT 12

US-09-071-035-114  
; Sequence 114, Application US/09071035



Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 114:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 406 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-071-035-114

Query Match 4.3%; Score 106; DB 4; Length 406;  
Best Local Similarity 23.0%; Pred. No. 0.023;  
Matches 102; Conservative 66; Mismatches 170; Indels 106; Gaps 25;  
QY 33 MKAATLMASMLRVDDLRGTGTVVLDPGAGVGSILTAALVDRLHTRPDVAVHV-----AV 86  
Db 1 MKKLKMMGIMLFV-----STVLV-----GCGTTATXKIDEKATEKTSVSKKVLNLMENSEI 51  
QY 87 ETDPFVVPVYLRTALEECRNAYGISYDLVEGDYLLNQAK-----LDGPFDLVIANPPYG 140  
Db 52 GMSDSIFTQDEASINAQSNVFEGLYQLDEKQDLIPAAAKEMPEISEDGK-RYTIKLRDGG 110  
QY 141 KLASDLSLARLATTARAVDPNVVYAFVWRAVISLKEQGRGVFVIVPSRWANGPYRQFRHW 200  
Db 111 KWSNGD-----AVTA-----NDVFPAW-RKLANPKNQANYFFLEGTILNG-----150  
QY 201 LMTAVS-----LDLHVFESRTKVFDATKVKQENVIVAFS-VRPQSSSVLSRSVAH 251  
Db 151 --TAITKEKAPBELGVKALDDYTLVLETKPVYFTSLIAFPFPQNEAFVKEKGQAY 208  
QY 252 G--EESIASVPF-----SALVHD-----EDDDKI-----VHF-----AESASVPSAAR 288  
Db 209 GTSSEMIVSNGPFLMKWQDSAMSWDFVRNPYYDKVKSETIHFELVKETTIVNLT-- 266  
QY 289 FTLADLIGIGVSTGKVDFF--RNRQYLTNDLDASGVV-PMVYQSNIRSGKIDWQVGA--- 342  
Db 267 YESGELDVAVLTG---DPAKQNR-----DNPDEAIEASRKVYSLRLNQKNEKPSIFANEN 319  
QY 343 -RKPQGFVAVEDVALRQLLPQGS-----YV-----VVKQTAKEDRRRVAIVWDGASRVA 392  
Db 320 VRKALAYALDKKSLVDNILDGSKSEIYIGYPEKVFVNPETNEDFRQAGALVKTDAKKA- 378  
QY 393 LDNKTNYLHESQRPLEKNVARGLM 416

Db 379 ----KEYLDKAKAELNGDVAILL 398  
RESULT 13  
US-09-134-000C-4354  
; Sequence 4354, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4354  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-4354  
Query Match 4.3%; Score 105; DB 4; Length 557;  
Best Local Similarity 23.0%; Pred. No. 0.05;  
Matches 102; Conservative 66; Mismatches 170; Indels 106; Gaps 25;  
QY 33 MKAATLMASMLRVDDLRGTGTVVLDPGAGVGSILTAALVDRLHTRPDVAVHV-----AV 86  
Db 8 MKKLKMMGIMLFV-----STVLV-----GCGTTATXKIDEKATEKTSVSKKVLNLMENSEI 58  
QY 87 ETDPFVVPVYLRTALEECRNAYGISYDLVEGDYLLNQAK-----LDGPFDLVIANPPYG 140  
Db 59 GMSDSIFTQDEASINAQSNVFEGLYQLDEKQDLIPAAAKEMPEISEDGK-RYTIKLRDGG 117  
QY 141 KLASDLSLARLATTARAVDPNVVYAFVWRAVISLKEQGRGVFVIVPSRWANGPYRQFRHW 200  
Db 118 KWSNGD-----AVTA-----NDVFPAW-RKLANPKNQANYFFLEGTILNG-----157  
QY 201 LMTAVS-----LDLHVFESRTKVFDATKVKQENVIVAFS-VRPQSSSVLSRSVAH 251  
Db 158 --TAITKEKAPBELGVKALDDYTLVLETKPVYFTSLIAFPFPQNEAFVKEKGQAY 215  
QY 252 G--EESIASVPF-----SALVHD-----EDDDKI-----VHF-----AESASVPSAAR 288  
Db 216 GTSSEMIVSNGPFLMKWQDSAMSWDFVRNPYYDKVKSETIHFELVKETTIVNLT-- 273  
QY 289 FTLADLIGIGVSTGKVDFF--RNRQYLTNDLDASGVV-PMVYQSNIRSGKIDWQVGA--- 342  
Db 274 YESGELDVAVLTG---DPAKQNR-----DNPDEAIEASRKVYSLRLNQKNEKPSIFANEN 326  
QY 343 -RKPQGFVAVEDVALRQLLPQGS-----YV-----VVKQTAKEDRRRVAIVWDGASRVA 392  
Db 327 VRKALAYALDKKSLVDNILDGSKSEIYIGYPEKVFVNPETNEDFRQAGALVKTDAKKA- 385  
QY 393 LDNKTNYLHESQRPLEKNVARGLM 416  
Db 386 ----KEYLDKAKAELNGDVAILL 405

RESULT 14  
US-09-489-039A-10635  
; Sequence 10635, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709,2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27

US-08-311-731A-13  
; Sequence 13, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311,731A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: C0044/7125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 202 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: October 1, 2004, 16:13:07 ; Search time 103.928 Seconds  
(without alignments)  
1504.830 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHPASTETRRQAALGKLDP.....QETLDVAVARLCEIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	223	9.0	1068	10	US-09-966-997-11
2	164	6.6	543	12	US-10-335-977-7269
3	161	6.5	543	12	US-10-335-977-7268
4	141	5.7	571	12	US-10-282-122A-51640
5	135	5.5	514	12	US-10-282-122A-55227
6	134	5.4	514	12	US-10-282-122A-65854
7	134	5.4	534	12	US-10-282-122A-72313
8	133.5	5.4	381	12	US-10-335-977-7481
9	133.5	5.4	384	12	US-10-335-977-7482
10	130	5.3	545	10	US-09-882-227-600
11	126	5.1	545	12	US-10-335-977-7271
12	125.5	5.1	317	9	US-09-712-363-185
13	125.5	5.1	317	12	US-10-282-122A-62340
14	125.5	5.1	317	12	US-10-282-122A-64484
15	124.5	5.0	295	12	US-10-282-122A-53706

Sequence 10737, A  
Sequence 56854, A  
Sequence 7395, Ap  
Sequence 7399, Ap  
Sequence 164, App  
Sequence 57917, A  
Sequence 8, Appli  
Sequence 7396, Ap  
Sequence 76713, A  
Sequence 244, App  
Sequence 72248, A  
Sequence 44252, A  
Sequence 7272, Ap  
Sequence 13654, A  
Sequence 116, App  
Sequence 74193, A  
Sequence 7411, Ap  
Sequence 7412, Ap  
Sequence 48561, A  
Sequence 12, Appl  
Sequence 58541, A  
Sequence 44255, A  
Sequence 13182, A  
Sequence 74177, A  
Sequence 6, Appli  
Sequence 3277, Ap  
Sequence 74345, A  
Sequence 12279, A  
Sequence 44253, A  
Sequence 70377, A

16 124.5 5.0 511 9 US-09-815-242-10737  
17 124.5 5.0 530 12 US-10-282-122A-56854  
18 122.5 5.0 491 12 US-10-335-977-7395  
19 122.5 5.0 1253 12 US-10-335-977-7399  
20 122 4.9 358 10 US-09-882-227-164  
21 122 4.9 530 12 US-10-282-122A-57917  
22 121.5 4.9 537 9 US-09-486-734A-8  
23 121.5 4.9 591 12 US-10-335-977-7396  
24 120 4.9 348 12 US-10-282-122A-76713  
25 119 4.8 823 9 US-09-895-913A-244  
26 118 4.8 291 12 US-10-282-122A-72248  
27 117 4.7 518 12 US-10-282-122A-44252  
28 117 4.7 550 12 US-10-335-977-7272  
29 116 4.7 317 9 US-09-815-242-13654  
30 116 4.7 317 10 US-09-769-744A-116  
31 116 4.7 317 12 US-10-282-122A-74193  
32 116 4.7 815 12 US-10-335-977-7411  
33 116 4.7 842 12 US-10-335-977-7412  
34 115.5 4.7 530 12 US-10-282-122A-48561  
35 115.5 4.7 959 16 US-10-616-624-12  
36 115 4.7 292 12 US-10-282-122A-58541  
37 115 4.7 518 12 US-10-282-122A-44255  
38 114 4.6 290 9 US-09-815-242-13182  
39 114 4.6 290 12 US-10-282-122A-74177  
40 113.5 4.6 515 9 US-09-486-734A-6  
41 113.5 4.6 1026 15 US-10-369-493-3277  
42 113 4.6 317 12 US-10-282-122A-74345  
43 113 4.6 518 9 US-09-815-242-12279  
44 113 4.6 518 12 US-10-282-122A-44253  
45 113 4.6 518 12 US-10-282-122A-70377

## ALIGNMENTS

### RESULT 1

US-09-966-997-11  
; Sequence 11, Application US/09966997  
; Publication No. US20030100052A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Shuang-yong  
; APPLICANT: Dore, Andrew  
; APPLICANT: Hume, Adam  
; APPLICANT: Pelletier, John  
; APPLICANT: Zhou, Jing  
; TITLE OF INVENTION: Method For Cloning And Expression of BsmBI Restriction  
; TITLE OF INVENTION: Endonuclease And BsmBI Methylase In E. coli And  
; TITLE OF INVENTION: Purification Of BsmBI Endonuclease  
; FILE REFERENCE: NEB-191  
; CURRENT APPLICATION NUMBER: US/09/966,997  
; CURRENT FILING DATE: 2001-09-28  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 1068  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus B61  
US-09-966-997-11

Query Match 9.0%; Score 223; DB 10; Length 1068;  
Best Local Similarity 20.7%; Pred. No. 1.9e-12;  
Matches 112; Conservative 86; Mismatches 219; Indels 124; Gaps 21;  
QY 21 TTAQVLGQFFTPMKAATLMAFMLRVDDLRGTVRVLPFGAGVGSLLTA-ALVDRLHTRPDV 79  
118 TASKAVGYFFTPISLGRVMVKLAKDKPKNLKSIQVDPACGIGSLALALI-----YNPEI 172  
QY 80 AVHVAVEDTDFVVPVYLRATLEECRNAYGIS--YDLVEG---DYLLNQAKLDPGFDLVI 134  
173 E-NVVGIELDSFTANISHKLLVRISKDLGTPKIKIINQNFLLVYLNVEEHEKFDLLI 231  
QY 135 ANPPYCKL-----ASDSLRLATTARAVDVP 160

Db 232 MNPYGRVRLKNSLTNETKSGLTGISELEKRLREETILMAADLRKKFASVGLKGTG 291  
QY 161 NVYVAFWRAVISLKEQGRGVFIVPRSWANGPYYPQFRHMLMTAVSLDILHVFESRTKVF 220  
Db 292 EYSKFLAISTKIVKONGVIAITPSSWLGDESGRELKYLIVENHGISCINWFKESAKLF 351  
QY 221 ADTKQENVIIVAFSVRPOSSVWLSRVSVAHGEESIASVSPSALVHDEDDDKIVHFAES 280  
Db 352 SG--VNQPTTVKIKVNSKIEIQG-----PLSSL---EELGRDQYLD 393  
QY 281 ASVP--SAARFTLADLGIG-----VSTGKVVDPNRQYITDNDASGVVPM 324  
Db 394 CNIKKYSPEWYRIPQCGNERAKILSKLHNHAPLSHKKIYNLRGELDTLTKHLLSDNPN 453  
QY 325 VQSNIRSGKIDWPQVARKPQ-----GFVAVEDVAL-----RQLPQ 362  
Db 454 HWRL--IRGDHVE--KFNLKNPEESEKLGIFV---DHQLFKRMGKSNKLRHKNWRITLPQ 507  
QY 363 GSYVVVKQTAKEDRRRVIAAVWDGASVALDNKTNYLH--ESQRPLEKNVARGLMWLNS 421  
Db 508 CSYM-----NKKRIEACIVEPNNIIA--NSCNYITLEDGNELVNLL--LCAINS 556  
QY 422 TVLDQYFRAFSGHTOVNAGDLRLPLC-----REDLILAKVVPDGLPDQETLDVAVRL 477  
Db 557 AVIEWRFRFLNSNNHVSVEIDEFFIFKFDTEMTLMLKGLFKHPLENWSKIEALIALM 616  
QY 478 F 478  
Db 617 Y 617

## RESULT 2

US-10-335-977-7269  
; Sequence 7269, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:

; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 10031  
; CORRESPONDENCE ADDRESS:

; ADDRESSER: LAHIVE & COCKFIELD  
; STREET: 28 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: Windows NT 4.0  
; SOFTWARE: UNIX

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/335,977  
; FILING DATE: 30-Dec-2002

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/993,002

; FILING DATE: 17-DEC-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Mandragouras, Amy E.

; REGISTRATION NUMBER: 36,207

; REFERENCE/DOCKET NUMBER: GTN-018

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)742-4214

; INFORMATION FOR SEQ ID NO: 7269:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 543 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...543  
; SEQUENCE DESCRIPTION: SEQ ID NO: 7269:  
US-10-335-977-7269

Query Match 6.6%; Score 164; DB 12; Length 543;

Best Local Similarity 23.3%; Pred No. 5.2e-07;  
Matches 78; Conservative 61; Mismatches 124; Indels 72; Gaps 16;

QY 28 QFFPMKATLMASMLRVDDLGRTVRLDPGAGVSGSLTAALVDRJLHTRPDPVAVHVAVE 87

Db 208 EYVTFSLASIIAKLL-VNKPTQSVKIYDPSAGTGTLLMALAHQIGTD--SCTLYAQDIS 264

QY 88 TDPFVVPVIRATLECRNAYGISYDL---VEGDYLLNQAKLD--GPFDLVIANPPYGL 142

Db 265 QKSLRMLKLNILND-----LTHSLRHAIEGNTLTNPYHSHKDHKGKMDFIYNSPPFKLD 318

QY 143 ASDSLARLATTAR--AVDPVN-----VYVAFWRAVISLKEQGRGVFIVPR---S 187

Db 319 FSNEHAEISQNKDNDFLGVNIPKNDKSKMPIYTLFFOHLNMLSPKKGALIVTGFIS 378

QY 188 WANGPYYPQFRHMLTAVSLDILHVFESRTKVFAADTKVKQENV-IVAFSVRPOSSSVLUS 246

Db 379 AKSGVNNKNNVRLHVDRL---VYGVICMPSQVFANTGT---NVSIIFFOKTSPSAKEVILI 432

QY 247 RSVAHGEE-----SIASSVPFSALV-HDEDDDKIVHFAESA 281

Db 433 DASKLGEETENKNNKTRLPSPDMDLILETFONKAPKSDFCALVSFDEITEK-----NY 486

QY 282 SVPSAARFTLADLGIGVSTGKVVDPRN--ROYLTD 314

Db 487 SLNPOQYFTIEDTSETISQ---AEFENLMQYISSE 518

## RESULT 3

US-10-335-977-7268  
; Sequence 7268, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:

; APPLICANT: DOUGLAS SMITH et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 10031

; CORRESPONDENCE ADDRESS:

; ADDRESSER: LAHIVE & COCKFIELD

; STREET: 28 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: Windows NT 4.0

; SOFTWARE: UNIX

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/335,977

; FILING DATE: 30-Dec-2002

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/993,002

; FILING DATE: 17-DEC-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Mandragouras, Amy E.

; REGISTRATION NUMBER: 36,207

; REFERENCE/DOCKET NUMBER: GTN-018

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)742-4214

; INFORMATION FOR SEQ ID NO: 7268:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...543
; SEQUENCE DESCRIPTION: SEQ ID NO: 7268:
US-10-335-977-7268

Query Match 6.5%; Score 161; DB 12; Length 543;
Best Local Similarity 23.3%; Pred. No. 1e-06;
Matches 78; Conservative 61; Mismatches 124; Indels 72; Gaps 16;

QY 28 QFTTPMKAATLMASMLRVDLRTGTVRVLDPGAGVGSILTAALVDRLHTRPDVAHVHVAVE 87
Db 208 EYITPLSIASIIIAKL-VNKPTQSVKIYDPSAGTGLMALAHQIGTD--SCTLYAQDIS 264
QY 88 TDFVVPYLRATLEECRNAYGISYDL---VEGYLLNQAKLD--GPFDLVIANPPYKGL 142
Db 265 QKSLRMLKLNLIIND-----LTHSLRHAIEGNTLTNPYHSKDHKGMDFIVSNPPFKLD 318
QY 143 ADSLRLATLATTAR--AVDPVN-----VYAFWVRAVISLKEQGRGVFIVPR---S 187
Db 319 FSNHAEIAISONKNDFFLGVFNIPNDKSKMPIYTLFFQHCLNMLSPKKGAIIVPTGTFIS 378
QY 188 WANGPYVROFRHMLTAVSLDLHVFESRTKVPADTKVKQENV-IVAFSVRPOSSSVLS 246
Db 379 AKSGVENKIVRHLDVRL---VYGVICMPSQVPANTGT---NVSIIFFQKTPSAKEVILI 432
QY 247 RSVAHGEE-----SIASSVPFSAIV-HDEDDDKIVHFAESA 281
Db 433 DASKLGEYTENKNNKTRLRPSMDLILETFONKAPKSPDFCALVSFDEITEK-----NY 486
QY 282 SVPSAARFTLADIGIGVSTGKVVDVFN--ROYLTD 314
Db 487 SLNPGQVFTIEDTSETISQ---AEFENLMOQYSSE 518

RESULT 4
US-10-282-122A-51640
; Sequence 51640, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

```

```

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51640
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51640

Query Match 5.7%; Score 141; DB 12; Length 571;
Best Local Similarity 19.7%; Pred. No. 0.00011;
Matches 112; Conservative 85; Mismatches 213; Indels 158; Gaps 27;

QY 19 DPTTQAVLQGFPTPKAATLMASMLRVDLRTGTVRVLDPGAGVGSILTAALVDRLHTRPD 78
Db 6 DATEERLTGRYFTPDRLASYIIDWIQDN--NINKILEPSCGNGVFLECLGER----RLE 59
QY 79 VAVHVVAVTDTDFVYPYLRATLE---ECRNAYGISYDLVEGYLLNQ----- 123
Db 60 DGRNITAIDEIDV--SFEASMQIDNSLRFNFCYQALRNDGDIVNNGIVINDDFYK 117
QY 124 ---AKLDGP-FDLVIANPPY-----GKLASDSLRLATLATTARAVDVNPNVAVFVRAV 171
Db 118 VVEQLQGRQFQAIQVNGNPPYRYQYISEQOREBQSKILIRNNRPNKILINAWFSFVACA 177
QY 172 ISLKEQGRGVFIVPRSWANGPYRQFRHMLTAVSLDLHVFESRTKVPADTKVKQENVI 231
Db 178 EILDGNGKGLVIPAELLQVAYAEIDLRRFMRTFQRTITVTF--RELVPFN--VQEEVVL 233
QY 232 V-----AFSVRPOSSSVLSRSVAH--GEESIASVPFSAIVHDE----- 269
Db 234 LLVEKEILHTRHQRLRIVEYQDINELTESNDLDEYFPFNDVEINESKWTYFLSANDIRLI 293
QY 270 ----DDDKIVHFAESASVPSAARFTLADIGIGVSTGKVVDVFNROYLTONLDASGVV-PM 324
Db 294 NNIRENDKFFRFSDFARVE-----VGITGNNDYFCVNRVRVEYDILERVCRPL 342
QY 325 V-----YQSNIRSGK-----IDWPQVGARKPQGFVAVEDVALRQLLPQGSY 365
Db 343 IARSVNNGVVKFTYEDWKSNIENGAKTYLIDFPDPYDEVQ-----LSYKQVIEYGIH 395
QY 366 VVVKROTAK---EDRRRVIAAVW-----DCASRVLD 394
Db 396 -NDKNSGYKCRIRDRWYKVPESIWVPDAFFLRRNHLXPKFVINSQEVNAVSTDMHRVFN 454
QY 395 NKTNYLHESQRPLEKNVARGMLMLNLTVDQYFRAFSGHTQVNAAGDLRLPFLCREDLI 454
Db 455 N-----HDDRE-----RILSYNSIAL--AFTEIEGRS-YGGGVLEILP---RE--- 493
QY 455 LIAKVVDPGLPDQETL-DAVARLFCBI 481
Db 494 -VGRIMPDLNRRELIDDATVSNLLNRI 520

RESULT 5
US-10-282-122A-65227
; Sequence 65227, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl

```

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65227
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-65227

Query Match      5.4%; Score 135; DB 12; Length 514;
Best Local Similarity 22.0%; Pred. No. 0.00038;
Matches 76; Conservative 52; Mismatches 128; Indels 90; Gaps 15;

Qy 27 QGFPTPKAATLMASMLRVDDLRTGTVRVLDPGAGVGSLSLAALVDRHLHTERPDVAHVAV 86
Db 196 GEFTPOSVSKLIA-RLAVHGOEKVKNKYDPACGSGSL-----LLOAKKQFDEHI--I 245
Qy 87 ETDPPVVPYLRATLEECRN---AYGISYD---LVEGDYLLNOGAKLDGPFDLVIANPPYG 140
Db 246 BEGFGQEIHTTNYLARMNFMFLNVNKNKHELGDTLTNPKLKDSKPFDAVSNPPYS 305
Qy 141 -----KLASDLARLATTARAVDPNVVYAFVWRAVISLKEQGRGVFI-VPRSWAN 190
Db 306 IDWIGSDPTLIND--RFAPAGVLAPSKADFAFILHALNVLSGRGAIVSPFGIFYR 363
Qy 191 GPYRQRFHWM-----TAVSLDILHVFESTRTKVFADTKVQENVIV 232
Db 364 GGAEQKIROYLVEGNYVETVIALAPNLPFYGTIAVNIIVLSKHK-----DNTDI 412
Qy 233 AF-----SVRPQSSSVLSRSVAHGEESIASVPFSAVHDEDDDKIVHPAESASVP-- 284
Db 413 QFIDAGGFFKETTNNVLT-----BEHIAEIVKL-----FADKADVPHIA 452
Qy 285 -SAARFTLADIGVSTGKVDFFNRQYLT-----NLDASGVVPMV 325
Db 453 QNAAQQTVDKNGYNLAVSYVEAETREVIDIRQINAEISVAKI 498

RESULT 6
US-10-282-122A-65854
```

```
; Sequence 65854, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65854
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-282-122A-65854

Query Match      5.4%; Score 134; DB 12; Length 514;
Best Local Similarity 22.0%; Pred. No. 0.00048;
Matches 76; Conservative 52; Mismatches 128; Indels 90; Gaps 15;

Qy 27 QGFPTPKAATLMASMLRVDDLRTGTVRVLDPGAGVGSLSLAALVDRHLHTERPDVAHVAV 86
Db 196 GEFTPOSVSKLIA-RLAVHGOEKVKNKYDPACGSGSL-----LLOAKKQFDEHI--I 245
Qy 87 ETDPPVVPYLRATLEECRN---AYGISYD---LVEGDYLLNOGAKLDGPFDLVIANPPYG 140
Db 246 BEGFGQEIHTTNYLARMNFMFLNVNKNKHELGDTLTNPKLKDSKPFDAVSNPPYS 305
Qy 141 -----KLASDLARLATTARAVDPNVVYAFVWRAVISLKEQGRGVFI-VPRSWAN 190
Db 306 INWIGSDPTLIND--RFAPAGVLAPSKADFAFILHALNVLSGRGAIVSPFGIFYR 363
Qy 191 GPYRQRFHWM-----TAVSLDILHVFESTRTKVFADTKVQENVIV 232
Db 364 GGAEQKIROYLVEGNYVETVIALAPNLPFYGTIAVNIIVLSKHK-----DNTDI 412
Qy 233 AF-----SVRPQSSSVLSRSVAHGEESIASVPFSAVHDEDDDKIVHPAESASVP-- 284
Db 413 QFIDAGGFFKETTNNVLT-----BEHIAEIVKL-----FADKADVPHIA 452
Qy 285 -SAARFTLADIGVSTGKVDFFNRQYLT-----NLDASGVVPMV 325
```

Db 453 QNAACQTVKDNQNGYNLAVSSVVEPEDTREIIDI KOLNAEISETVAKI 498

## RESULT 7

```

US-10-282-122A-72313
; Sequence 72313, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72313
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72313

```

Query Match	5.4%;	Score 134;	DB 12;	Length 534;
Best Local Similarity	20.9%;	Pred. No. 0.00051;		
Matches	62;	Conservative 56;	Mismatches 147;	Indels 32; Gaps 111;
Qy	27	GQFPTMKAAATLMASMLRVDDLGRGTVRVLDGAGVGSLSAALVDRLHTERPDVAHVHVAV	86	
Db	203	GEFTPTQVSKLLAQLVWVGK-EHINKVYDPTCSGSLLLQMKQKFETHILEEFGFGQEI	261	
Qy	87	ETDFPVVYLRATILEECRNAYGISVDLVEGDYLLNQAGKLDGPDFDLVIANPYG----	140	
Db	262	NMTNYNLARMNMFL---HNIYNPNFDIRRGDTLLNPQHLVYRPPDAIVSNPNYSYIKWIGD	318	
Qy	141	---KLASDLSRLATTAARAVDPNVVYVFWRAVLSLKEQGRGVFI-VPRSWANGPPYRQ	196	
Db	319	ADPTLLINDE--RFAPAGKLAPKSKADPAFTMHSLSHLNKGRAAIVCFPGIFYRGAECT	376	
Qy	197	FRHLWTAVSJDILHVFESRTKVFADTKVQENVIASFVSPQSSSVL-----SRSSVAH	251	
Db	377	IROYLDNNFVEA--VIALPNLNFSGTISATYVLVLAKN-KPEDTKLFDASSDEKSTVS	433	

QY	252	GEESIASVPFSAVHDEDDDKIVH-FAESA\$VPSAARFT-----LA---DLGIGVST 300
		:   : : :         :   :   :
Dp	434	GKKEFYETNGKITNPKNTEALVELFKKKKDVDYRAKLVDNKLIAENDYNLSVST 490

## RESULT 8

```

US-10-335-977-7481
; Sequence 7481, Application US/10335977
; Publication No. US2004005279A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7481:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...381
; SEQUENCE DESCRIPTION: SEQ ID NO: 7481:
US-10-335-977-7481

Query Match 5.4%; Score 133.5; DB 12; Length 381;
Best Local Similarity 26.6%; Pred.No. 0.00033;
Matches 54; Conservative 29; Mismatches 73; Indels 47; Gaps 10;

Qy 26 LGQFFETPMKAATLMASMLRVDDLGGTVRLVDPGAGVGSLSLTAALVDRLHTRPDVAHVHA 85
Db 13 LGQVFTP---KNIVDFMLILKHNHGS--VLEPSAGDGSFLKRL-----KKAAG 55
Qy 86 VETDPFVVPVYRATLBECRNAYGISYDLVBEGDYLLNQAKLDGPFDLIVIANPPYK---L 142
Db 56 IEIDPKICP-----KNALCMDF----FDY-----PLENQFNIIIGNPFVVKHDI 96
Qy 143 ASDSLARLATTARVDVNVVAFWTRAVLSLKEQRCGVFIVPRSWANGPYVQFRHWM 202
Db 97 APSTKEXKLHSL-FDERSNLVLEFIEKAIKHLKPKGELIFITPRDFUKSTSSVKLENIW 155
Ov 203 TAVSLDILHVFEGRTKVFADTKV 225

```



Db 156 KEGT--ITHFFE-----LGDQKI 171

## RESULT 9

US-10-335-977-7482  
; Sequence 7482, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-DEC-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragoras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 7482:

SEQUENCE CHARACTERISTICS:

LENGTH: 384 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...384

SEQUENCE DESCRIPTION: SEQ ID NO: 7482:

US-10-335-977-7482

Query Match 5.4%; Score 133.5; DB 12; Length 384;  
Best Local Similarity 26.6%; Pred. No. 0.00034;  
Matches 54; Conservative 29; Mismatches 73; Indels 47; Gaps 10;

QY 26 LGQFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVSLTAALVDRLHTERPDVAHVVA 85

Db 16 LGQVTFP---KNIVDFMLTKENHGS--VLEPSAGDSFLKRL-----KKAVG 58

QY 86 VETDPFVVPYLRATLEECNAGVIGISVDLVEGYSLLNQAKLDGPFDLVIANPPYCK--L 142

Db 59 IEIDPKICP-----KNALCMDF----FDY-----PLENQFTIIGNPPYVKKHDI 99

QY 143 ASDSLARLATTARAVDPNVYVAFWVRVAVISLKEQGRGVFIYRPSWANGPYVQRPHMLM 202

Db 100 APSTKEKLHSL-FDERSNLVLFETKAIKLPKXGELIFITPDRFLKSTSSVKLNWIY 158

QY 203 TAVSLDLHVFESRTKVPADTKV 225

Db 159 KEGT--ITHFFE-----LGDQKI 174

## RESULT 10

US-09-882-227-600

; Sequence 600, Application US/09882227

; Publication No. US20030158396A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Oomenen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20030158396A1el Helicobacter Polypeptides in the

; TITLE OF INVENTION: Genome

; FILE REFERENCE: 06132/047002

; CURRENT APPLICATION NUMBER: US/09/882,227

; CURRENT FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 08/902,615

; PRIOR FILING DATE: 1997-07-29

; NUMBER OF SEQ ID NOS: 638

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 600

; LENGTH: 545

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-882-227-600

Query Match 5.3%; Score 130; DB 10; Length 545;

Best Local Similarity 18.9%; Pred. No. 0.0013;

Matches 86; Conservative 87; Mismatches 203; Indels 78; Gaps 19;

QY 17 KLDPTTQAVLGQFTPEMKAATLMASMLRVDDLRTGTVRVLDPGAGVSLTAALVDRLHTER 76

Db 101 ELSNATRNLEGYYTPNRIVEQLFTLPKDFVDSQAI-FCDDPAVSGNFI-----MHALK 153

QY 77 PDVAV-HVVAVETDPFVVPYLRATLEE-----CRNAYGISVDLVEGDYLLNQGAKLDGPF 130

Db 154 LGFKVENIYGYDTDAFAVALTKKIKERYHLDCLN-----IVQKDFL---NLKHTPOF 203

QY 131 DLVIANPPYKGL-----ASDSLARLATTARAVDPNVYVAFWVRVAVISLKEQGRGVFIYPR 186

Db 204 DCIFTNPPWGGKYNQNKFNKQFNLSQSLDSASL---FFTIASLNCLENKNAHLGLLPE 260

QY 187 SWANGPYVQRPHWMLTAVSLDLHVFESRTKVPADTKVQ---ENVIVAFVSRP---QS 240

Db 261 SCLNIDAFKKVREMAK-----FHIRSLIDFKPKNLTAKVGLAKKTINKOOK 311

QY 241 SSVLSRSVAHGEESIASVSP---FSALVHDEDDDKIVHFAESASVPSAARFTLADLGIG 297

Db 312 ISCFYQNSKFKRSPSSFFNNPKKIFNIHCSKENKILDHL---FSLPHMTLKNNAHFALG 368

QY 298 VSTGKVVDFRQRYLTDLNDSAGVVPVYQNSIRSGKIDWPOVGARKPGQFVAVEDVALR 357

Db 369 IVTG-----NNKEKLHPKQE-KNTPIPRGSDILK-----GLKAPSQFTNAGLKDCQ 415

QY 358 QLLPGQSY-----VVVKQTKAKEDRRRVIAAVWDGASRVALDNKNTYLHESORPLEXNVA 412

Db 416 QVAPLSLQAREKIVYKFISSK-----LVFFYDNKQRLFLNSANFVLKENFPINAHAL 469

QY 413 RGLMLWMLNSTVLVDQYFRAPFSGHTQVNAAGDLRLRP 446

Db 470 KEL---LNSDLMQFTFESLFTKILRKDLECLP 500

## RESULT 11

US-10-335-977-7271

; Sequence 7271, Application US/10335977

; Publication No. US20040052799A1

; GENERAL INFORMATION:

; APPLICANT: DOUGLAS SMITH et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

; RELATING TO HELICOBACTER PYLORI FOR

## DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GIN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 7271:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 545 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...545  
SEQUENCE DESCRIPTION: SEQ ID NO: 7271:  
US-10-335-977-7271

Query Match	5.1%;	Score 126;	DB 12;	Length 545;
Best Local Similarity	17.5%;	Pred. No. 0.0034;		
Matches	83;	Conservative	96;	Mismatches 185; Indels 110; Gaps 21;
Qy	17	KLDDTTQAVLGQFFTPMKAATMASLMRLVDDLRGTVRVLDPGAGVSGLSLTAALVDRLATER	76	
Db	101	ELSNTRNLGSIYTPRIIVEQLFTLPKDFTTQAI-FCDPVCSGNFV-----MHALK	153	
Qy	77	PDVAV-HVVAVETDPFVPVYLRLTEE-----CRNAYGISVDLVEGDYLLNQGAKLGGPF	130	
Db	154	LGPKVENIYGYDTDAFAVALTKKRIKERYLDCFN-----IMKQDFL-----SLKHAPQF	203	
Qy	131	DLVTANPYGKL-----ASDSLRLATTARAVDVPNVYVAFVWRAVISLKEQGRGVFIWPR	186	
Db	204	DCIETNPPGWKKYNONOKENFKQFNHLQSLDSASL-----FFMASLNCLEKWAYLGLLLE	260	
Qy	187	SWANGPY-----RQFRHMLMTAVSLDILHVFESERTKV-----FADTK	224	
Db	261	SLCLNIDAFSKMREALKFQMESLTDFDKPFKNLMTKAVGLAKKTHNKDQKISCFYQNSE	320	
Qy	225	VKQENVIAFSVRQSSSVLSRVAGEESIAS-----SVPSALVHDEDDDKIVHFAESA	281	
Db	321	FKRSP--SSFFNPNPK--I-FNIHCSHOKNKLHLFSLPHITLKN-----AHFA-----	366	
Qy	282	SVPSAARFTLADIGIVGTGKGVDFRPNRYQLTDNLDSGVVPMVYQNSIRGSKIDWPQVG	341	
Db	367	-----LGIVTG-----NNKEKHLPKQE-KNTTPIFRGSDILKDLR-----	400	
Qy	342	ARKPQGFVAVEDVALRQLLPQGSY-----VVVKRQTAKEDRRRRVITAAVWDGASRVALDNK	396	
Db	401	-KAPSOFTNADLKDCCOVAPLSLYOAREKIVVYKFLSSK-----LVFFYDNKORFLFNSA	453	

```

397  TNYLHESQRPLEKNVARGMLNLSNVLDQYFPAFSGHTQVNVAGDLRLPLCLR 450
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
454  NMFVLKENFPINAHKEL---LNSDLMOFIFESLFTKTHILRKDLCECLPLFAQ 504
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 12
US-09-712-363-185
; Sequence 185, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712.363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-185

```

	Query Match	5.1%; Score 125.5; DB 9; Length 317;	
	Best Local Similarity	25.6%, pred.No.0.0016;	
	Matches	73; Conservative 48; Mismatches 109; Indels 55; Gaps 13	
Qy	4	IASETTRQAALGKLDPPTQAVLQGFTTPMKAAATLMASMLRVDOLRGTVRVLDPGAGVGS	63
Dd	14	LGRTEIRRLAK--ELDFPRKSLGCNQF--VHDANTVRVVVAASGVSRSDLVLEVPGGHGS	69
Qy	64	LTAALVDLLHTERPDVAHVHVAVETDPPVVPYLRATILLECRCNAYGISVDLEGDVLLMQG	123
Dd	70	LTTALLDRGAT-----VTAVEIDPLAARLOQTVAEHSSEVHRITVNRRDLVALRR	121
Qy	124	AKLGPFDLVIANPYPKGLASDSLARLATATTARAVIDPNNV-VYAFAWRVAI SLK---EQGR	179
Dd	122	EDLAAPTAVVANULPY-NNAVPA LLHL----LVEFFSIRVVTVMVQAEVAERLAAEPGS	175
Qy	180	GVFIVPSRWANGPYFRQH-----WLMTAVSLDLIHV--FES-----RTKV	219
Dd	176	KEYGVPS--VKLREFGFRRRCMGVSPTFWPIPVYSGLRIDRYETSFPWPTDDAFRRRV	233
Qy	220	FADTKVQGENVIVAFSPVPOSSSVVLRSVAHGEE-----IASSV	260
Dd	234	F-----ELVDIAFAORCKTSRNAFVQAWAGSGGESANRLLAASI	271

RESULT 13  
US-10-282-122A-62340  
; Sequence 62340, Application US/10282122A

Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 62340

LENGTH: 317

TYPE: PRT

ORGANISM: Mycobacterium bovis

US-10-282-122A-62340

Query Match 5.1%; Score 125.5; DB 12; Length 317;  
Best Local Similarity 25.6%; Pred. No. 0.0016;  
Matches 73; Conservative 48; Mismatches 109; Indels 55; Gaps 13;

Qy 4 IASTETRRQALGKIDPTTQAVLGQFFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVGS 63  
Db 14 LGREIRRLAK--ELDFPRKSLQGNF--VHDANTVRVVAASGVSRSDLVLEVPGLGS 69  
Qy 64 LTAALVDRLHTERDPDVAVHVAVETDPFVVPYLRATLEECNAYGIGSYDLVEGDYLLNQ 123  
Db 70 LTLALLDRGAT-----VTAVEIDPLLASRLQQTVAEHSVSHVRLTVVNRDVLALRR 121  
Qy 124 AKLDGPPDLVIANPPYKGLASDSLARLATTARAVDENV-VYAFWRAVISLK---RQGR 179  
Db 122 EDLAAAPTAVANLUPY-NVAVPALLHL-----LVEFESIRVVTVMQAEVAERLAEPGS 175  
Qy 180 GVFTVPRSWANGPYRQFRH-----WMTAVSLDILHV--FES-----RTKV 219  
Db 176 KEYGVPS--VKLRFEGVRRCGMVSPVTFVPIRVYSGLVRIIDRYETSPWPTDDAFRRRV 233  
Qy 220 FADTKVQENVIYAFSVRPOSSSVLSRSVAHGEES-----IASSV 260  
Db 234 F-----ELVDIAFAQRKTSRNAFVQWAGSGSESANRLAASI 271

RESULT 14

US-10-282-122A-64484  
Sequence 64484, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:

APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 64484

LENGTH: 317

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-64484

Query Match 5.1%; Score 125.5; DB 12; Length 317;  
Best Local Similarity 25.6%; Pred. No. 0.0016;  
Matches 73; Conservative 48; Mismatches 109; Indels 55; Gaps 13;

Qy 4 IASTETRRQALGKIDPTTQAVLGQFFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVGS 63  
Db 14 LGREIRRLAK--ELDFPRKSLQGNF--VHDANTVRVVAASGVSRSDLVLEVPGLGS 69  
Qy 64 LTAALVDRLHTERDPDVAVHVAVETDPFVVPYLRATLEECNAYGIGSYDLVEGDYLLNQ 123  
Db 70 LTLALLDRGAT-----VTAVEIDPLLASRLQQTVAEHSVSHVRLTVVNRDVLALRR 121  
Qy 124 AKLDGPPDLVIANPPYKGLASDSLARLATTARAVDENV-VYAFWRAVISLK---RQGR 179  
Db 122 EDLAAAPTAVANLUPY-NVAVPALLHL-----LVEFESIRVVTVMQAEVAERLAEPGS 175  
Qy 180 GVFTVPRSWANGPYRQFRH-----WMTAVSLDILHV--FES-----RTKV 219  
Db 176 KEYGVPS--VKLRFEGVRRCGMVSPVTFVPIRVYSGLVRIIDRYETSPWPTDDAFRRRV 233  
Qy 220 FADTKVQENVIYAFSVRPOSSSVLSRSVAHGEES-----IASSV 260  
Db 234 F-----ELVDIAFAQRKTSRNAFVQWAGSGSESANRLAASI 271

Job time : 105.928 secs

```
RESULT 15
US-10-282-122A-53706
; Sequence 53706, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53706
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Corynebacterium diptheriae
US-10-282-122A-53706

Query Match          5.0%; Score 124.5; DB 12; Length 295;
Best Local Similarity 27.9%; Pred. No. 0.0018;
Matches 57; Conservative 25; Mismatches 81; Indels 41; Gaps 8;

QY      4  IASTERRQALGKLDPTTQAVIGQFTTPKKAATLMASMLRVDDLRCGTVTVLDPGAGVGS 63
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      10  LGPVEIRQLAE--KLDVTPTKLGQNF--VHDPNTVMIVSAADLNSDDHVEVGPGLGS 65

QY      64  LTAALVDRLHTRPDVAHVAVVETDPFVVPVPLRATLEECRNAYGISYDLVGEGLNLNQ 123
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      66  LTLALL-----DTAQKVAVEIDPRLAQQLPLTVAERAGQPADRLNIHKDALTVP 117

QY      124  AKLDGPFDLVIANPPYG-----KLASDSLARIATTA--RAVDVP 160
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      118  DDIDHTALV-ANLPYNVSPVLHLHLQIFPTIRRLVMVQAEVADRLAADPGNKVGV 176

QY      161  NVYVAFW--VRVVISLKEQGRGVF 182
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      177  SVKASFYGNVRRAGSI---GKNVF 197
```

**This Page Blank (uspto)**

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	39.6	4.1	1201	13	BX381961	BX381961 BX381961
C 2	39.2	4.0	1201	9	AL544900	AL544900 AL544900
C 3	38.6	4.0	808	14	CK126515	CK126515 AGENCOURT
C 4	38.6	4.0	539	9	AA788783	AA788783 ah31b05. s

## ALIGNMENTS



High quality sequence stop: 690.

# FEATURES

Location/Qualifiers  
1. 808  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7049194"  
/tissue\_type="whole body"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_ZGC\_10"  
/note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;  
Bulk tissue was collected from a whole adult individual  
from the Tuebingen strain. 1st strand cDNA was primed with  
a Not I - oligo(dT) primer, double-stranded cDNA was  
cloned into the Not I and EcoRV sites of pExpress-1.  
Library was size-selected for >1 kb fragments. A  
normalized version of this library is also available  
(NIH\_ZGC\_7). Library was constructed by Open Biosystems  
(Huntsville, AL)."

## ORIGIN

Query Match 4.0%; Score 38.8; DB 14; Length 808;  
Best Local Similarity 53.7%; Pred. No. 8.8;  
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 48 TCGCGGGCGTCTAAAGCGTTTGGTTTACGGCGCAACGATATAGCTCGTAGCGC 107  
Db 770 TCGGAGTGGCATCTTCATGCAGCTCGCGTTTCAGCTTTGCGATNGAGCTCCAGGTGGCA 711  
QY 108 TGTGACATTCCTCGGCTTTCGCGTTTGAAGCGGGAGATCGTGGTTGACTTCGACAC 167  
Db 710 CDTGTCCCTGACAGCTCGGATTCATTTTGAAGCTGCTGATCATCGGTTGAGATCAGCAA 651  
QY 168 TCCAGCGCTTGGCGTTTCAGAGATCAT 194  
Db 650 TCTCAGCTTGGTTTACGAGGTGAT 624

## RESULT 4

AA788783 539 bp mRNA linear EST 12-JAN-1999  
ah31b05 s1 Soares parathyroid tumor NBHPA Homo sapiens cDNA clone  
1240401 3', similar to gb:K01144 HLA CLASS II HISTOCOMPATIBILITY  
ANTIGEN, GAMMA CHAIN PRECURSOR (HUMAN);, mRNA sequence.  
AA788783  
AA788783.1 GI:2848903  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 539)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Insert length: 1264 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 285.  
Location/Qualifiers  
1. 539  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

# FEATURES

Location/Qualifiers  
1. 539  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

/clone="1240401"  
/tissue\_type="parathyroid tumor"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares parathyroid tumor NBHPA"  
/note="Organ: parathyroid gland; Vector: pT7T3D  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer  
(5'-TGTTACCAATCTGAAGTGGGCGCGCCACCAATTTTTTTTTTTTTTTT  
TTTTT-3'), double-stranded cDNA was size selected, ligated  
to Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT7T3  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

## ORIGIN

Query Match 4.0%; Score 38.6; DB 9; Length 539;  
Best Local Similarity 50.8%; Pred. No. 8;  
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 709 CATGACGAGTTCGTGGGTGGCTTTTCTTATGAGGCGCGTGAAGGTAAAGTTCGGTTT 768  
Db 260 CTTGGAGAGCTTCATGCGCAGGTTCTCCAGCTGCAGGTTCTGGAGGTGACTGTCAAGTT 319  
QY 769 GATGAGGAGCGGATCGCAGCGCTGCAAGAGCTATTGCTTACACCTTCAGCGGTCTAATT 828  
Db 320 GTCAAGCGCGCTGTGCTGTGTACAGGAATAAGCGGTTTGTGCTCTCAGCTAAGAGC 379  
QY 829 TTGTAACTGCTTTTCAAAATCGTGCAGTCGATCGTGCCTCCCTGAGCTGGCTGG 888  
Db 380 AAAGTTACAGGATTGAAAAACCTCTGTACAAGGCTCCCGGCTGCAATTGCTCTCGGG 439  
QY 889 G 889  
Db 440 G 440

## RESULT 5

BM402665 510 bp mRNA linear EST 01-JUL-2002  
SLA006G12.34711 An expressed sequence tag (EST) collection from the  
resurrection plant Selaginella lepidophylla Selaginella  
lepidophylla cDNA clone SLA006G12 5, mRNA sequence.  
BM402665  
BM402665.1 GI:21643881  
EST.  
Selaginella lepidophylla  
Selaginella lepidophylla  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;  
Selaginella.  
1 (bases 1 to 510)  
Turriaga, G. and Cushman, J.C.  
An expressed sequence tag (EST) collection from the resurrection  
plant Selaginella lepidophylla  
Unpublished (2002)  
Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: [icushman@unr.edu](mailto:icushman@unr.edu)  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 006 row: G column: 12  
Seq primer: T3 20mer



```

FEATURES
  source
    High quality sequence stop: 510.
    Location/Qualifiers
      1..510
        /organism="Selaginella lepidophylla"
        /mol_type="mRNA"
        /db_xref="taxon:59777"
        /clone="SLA006G12"
        /tissue_type="microphyll fronds undergoing desiccation for
        2.5 h"
        /dev_stage="adult"
        /clone_lib="An expressed sequence tag (EST) collection
        from the resurrection plant Selaginella lepidophylla"
        /note="Vector: Lambda Uni-Zap XR, Bluescript SK; Site 1:
        EcoRI; Site 2: XhoI; Library construction was performed
        according to manufacture's (Stratagene, Inc.) recommended
        protocol for the Lambda UniZapXR vector and cDNA synthesis
        kit."

ORIGIN
  Query Match      3.9%; Score 37.6; DB 12; Length 510;
  Best Local Similarity 50.0%; Pred. No. 15;
  Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 386 ATGGAACGGAAGGCTCGAGGAATCTCTGTAGTCTTCTTGTGATGAACATCGAAGCGG 445
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 50 AGCAACGCAAGGCTCAGGATGAGCATGAGGCTACGGCGGCGGAAATCCGGACGAGG 109
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 446 TTAAGCGCGAGCGGAAGCTCTCCAGGCTGCGATGATCTCTGTGATCTCCTGTGTGCG 505
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 506 AGGAATTTCTCTCTCGCTCGCGGCGAGATCCATTGCTGAAGAAGATGTCGAAGAGT 565
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 170 ACGAACATCACGAGGTTGAGGAGATGAAGAAACGGCTGAAGAAATGAAGAAGAGG 229
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 566 TTGTGCGG 573
DB |||||

RESULT 6
LOCUS
DEFINITION
  BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
  clone CSODI072YF05 3-PRIME, mRNA sequence.
ACCESSION
  BX381961
VERSION
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
REFERENCE
  1 (bases 1 to 1201)
AUTHORS
  Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE
  Full-length cDNA libraries and normalization
JOURNAL
  Unpublished (2001)
COMMENT
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  Library was constructed by Life Technologies, a division of
  Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
  http://fulllength.invitrogen.com/InvitrogenCorporation1600
  Faraday Avenue Genoscope sequence ID : CSODI072CC03NP1.
  Location/Qualifiers
    1..1201
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="CSODI072YF05"
      /tissue_type="PLACENTA COT 25-NORMALIZED"
      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
      /note="1st strand cDNA was primed with a NotI-oligo(dt)

FEATURES
  source
    primer. Five prime end enriched, double-strand cDNA was
    digested with Not I and cloned into the Not I and EcoR V
    sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
  Query Match      3.9%; Score 37.6; DB 13; Length 1201;
  Best Local Similarity 5.8%; Pred. No. 25;
  Matches 29; Conservative 136; Mismatches 33; Indels 1; Gaps 1;

QY 189 GATCATGACTGTGTCGGCGAGCATTTGGCCAAAGCGTACGCCACCGAAGTCGAAGA 248
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 TTTCGTAAGAAGACGCTTCGGCAGTGGTTCATATATGGCTTCGCGGTACTTAATCGGA 308
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 CCMNNNNKKKKCNKKKKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 465
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 309 CAATTAAACATCGCCACGAACTCGCAGCTCAACAGTACTGCTGTCTGACGAAGCAT 368
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 NNNTKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 525
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 ACAGCGCTAAGGCGATATGGAACGGAAGGCTT-CGAGGAATCTCTTGTAGTCTTCTTG 427
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 526 MNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 585
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 ATGAAGCATCGAAGCGGCTTAAAGCGGAGCGAGGAGCTCTCCAGGCTCGGATGCTCTG 487
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 586 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 645
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 488 TCGATCTCCCTGGTGGCGAGGAATTTCTGCTCTCGCTCGCGGCGAGAAATCCATGCTGA 547
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 646 KNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 705
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 548 AGAAGATGTCGAAGAGATTTCGCGGAGTTTGCACTCGCTCGCAGCGTCTCTACCTCG 607
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 706 NKNNKNTNKKKNNNAKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 765
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 608 GGGATATCTGTGGAAGCATTTCCCTATTTCGAACGAGAGATCTTTGAAGAGTGTCTCGCC 667
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 766 KNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 825
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
LOCUS
DEFINITION
  AA284745 488 bp mRNA linear EST 08-AUG-1997
  zt21g05.r1 Soares ovary tumor N5HOT Homo sapiens cDNA clone
  IWAG:713816 5', similar to Qb:K01144 HLA CLASS II
  HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN PRECURSOR (HUMAN); mRNA
  sequence.
ACCESSION
  AA284745
VERSION
  AA284745.1 GI:1927286
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
REFERENCE
  1 (bases 1 to 488)
AUTHORS
  Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
  Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
  Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
  Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
  Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
  Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
  and Marra, M.
TITLE
  Generation and analysis of 280,000 human expressed sequence tags
JOURNAL
  Genome Res. 6 (9), 807-828 (1996)
MEDLINE
  97044478
PUBMED
  8889549
COMMENT
  Contact: Wilson RK

```



```

QY 911 AAGATCCAGACCATCTGATTACCTTAACGG 941
| | | | | | | | | | | | | | | | | |
Db 378 AGGATCTCGCGATGAGTGGCTGCGCTTGGCGG 408

RESULT 10
BG907733 615 bp mRNA linear EST 05-JUN-2001
LOCUS Talr1162D06F Talr1 Triticum aestivum cDNA clone Talr1162D06 3',
DEFINITION mRNA sequence.
ACCESSION BG907733
VERSION BG907733.1 GI:14315409
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 615)
AUTHORS Cloutier, S., Dong, G. and Walsh, A.
TITLE Wheat functional genomics - Thatcher Lr1 cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.gc.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average inset size is >2.2 kb
Plate: 162 row: D column: 06
Seq primer: M13 Forward.
FEATURES
source
1..615
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="Talr1162D06"
/tissue_type="Leaf tissue"
/dev_stage="14 Days old"
/lab_host="E. coli XL0LR"
/clone_lib="Talr1"
/note="Vector: Lambda ZapII; mass excised in plasmid
vector pBK-CMV (Stratagene).; Site 1: EcoRI; Site 2: XhoI;
mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
inoculation with leaf rust pathogen Puccinia tritici
race BB5 carrying the avirulence gene Avr1."
ORIGIN
Query Match 3.8%; Score 36.6; DB 12; Length 615;
Best Local Similarity 62.6%; Pred. No. 33;
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 851 GTGAGTCGATCGTCATGGCTCCCTGAGTGGCTTGGGAACTGAGCGGTGGTAGCGG 910
| | | | | | | | | | | | | | | | | |
Db 301 GAGTGGGTGAGGCTTGCTGGTTCCCTTAGTATGTGACCCCTGACCGCTGGTAGCGG 360

QY 911 AAGATCCAGACCATCTGATTACCTTAACGG 941
| | | | | | | | | | | | | | | | | |
Db 361 AGGATCTCGCGATGAGTGGCTGCGCTTGGCGG 391

RESULT 11
CA623177/c 624 bp mRNA linear EST 23-NOV-2002
LOCUS wlin.pk0103.g10 wlin Triticum aestivum cDNA clone wlin.pk0103.g10
DEFINITION 5' end, mRNA sequence.
ACCESSION CA623177
VERSION CA623177.1 GI:25201473
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 624)
AUTHORS Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
Miao, G., Caraher, N. and Hanafey, M.K.
TITLE DuPont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
FEATURES
source
1..624
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wlin.pk0103.g10"
/tissue_type="leaf"
/clone_lib="wlin"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) leaf 7 day old
seedling, light grown (normalized)"
ORIGIN
Query Match 3.8%; Score 36.6; DB 14; Length 624;
Best Local Similarity 62.6%; Pred. No. 33;
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 851 GTGAGTCGATCGTCATGGCTCCCTGAGTGGCTTGGGAACTGAGCGGTGGTAGCGG 910
| | | | | | | | | | | | | | | | | |
Db 184 GAGTGGGTGAGGCTTGCTGGTTCCCTTAGTATGTGACCCCTGACCGCTGGTAGCGG 125

QY 911 AAGATCCAGACCATCTGATTACCTTAACGG 941
| | | | | | | | | | | | | | | | | |
Db 124 AGGATCTCGCGATGAGTGGCTGCGCTTGGCGG 94

RESULT 12
CB902040/c 761 bp mRNA linear EST 02-JUL-2003
LOCUS trico28xm20 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION jecorina cDNA clone trico28xm20, mRNA sequence.
ACCESSION CB902040
VERSION CB902040.1 GI:30116698
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE 1 (bases 1 to 761)
AUTHORS Foreman, P.K., Brown, D.E., Darkmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, E., Houfek, T.D., England, G.O.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
TITLE Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE 22803314
PUBMED 12788920
COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
FEATURES
Location/Qualifiers

```

```

source
1. 761
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric028xm20"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with Varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match      3.8%; Score 36.6; DB 14; Length 761;
Best Local Similarity 51.5%; Pred.No. 37;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 272 AGTGGGTGATATGGCTTCGCCGTACTTAATCGGACRAATTTAAACATCGCCACGAACT 331
Db 556 AGCCAGTAGCAATGCTCGCTGCCAGATCGCTCAAATTTGATAAATGACTGCATCT 497

QY 332 CGCAGCTCAACGAGTACTGCTGTCTGACGAAGCATTTACAGGGCGCTAAGGGCATATGGAA 391
Db 496 TGCAGGTTCAAGTCGTCATGTGTGTGATGCAGTACACGCCACAGGAGATGACA 437

QY 392 CGGAAGGCTTCGAGGAATCTTGTGTAGTCTTTCTTGATGAAGC 434
Db 436 AGGATCCCTGAGTTGCATCTGTGTTGAGAGCGTCCATTATGC 394

RESULT 13
CF871808/c
LOCUS
DEFINITION
CF871808
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1. (bases 1 to 761)
Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
Dean,R.A.
Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
Unpublished (2003)
Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: IT-F1 primer.
Location/Qualifiers
1. 761
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="tric028xm20"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/notes="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with Varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match      3.8%; Score 36.6; DB 14; Length 761;

```

```

1 359
source
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HC083905_r"
/clone_lib="Chlamydomonas reinhardtii 5% CO2"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"

```

	Query Match	3.7%;	Score 36.4;	DB 9;	Length 358;
	Best Local Similarity	61.7%;	Pred. No. 28;	Indels 0;	Gaps 0;
	Matches 58;	Conservative 0;	Mismatches 36;		
QY	403	GAGGATCTCTTGTAGTCTTTCTTTGATGAAGCATCGAAGCGCGTAAAGCGAGCGGAA	462		
Db	278	GGGGTACTCTTTGGGGTCTGCTTTGATGATGACGGCGCCCTTGGCGCCCTCGCGAGCGGAA	219		
QY	463	GCTCTCCAGCGTCGATGATCTCTGTGATCTCC	496		
Db	218	GCCTCTCTCTTGACGAAGCCGATGTAGATCTCC	185		

Search completed: October 2, 2004, 05:12:44  
Job time : 2828.71 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 19:25:51 ; Search time 3957.91 Seconds  
(without alignments)  
10644.353 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972  
Sequence: 1 grgaacagcagtgacggcat.....ttgggcgcgtacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_mam.\*

37: em.htg\_vrt.\*

38: em\_sy.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	156.6	16.1	343250	1	AP003594	AP003594 Nostoc sp
2	143	14.7	5039	1	YEA14030	YEA14030 Yersinia
3	135.8	14.0	2892	1	BACHSMR	LO1541 Bacillus su
4	58.8	6.0	3889	1	PROIRM	K02081 P.stuartii
5	45.4	4.7	300511	1	AE016775	AE016775 Pseudomon
6	38.6	4.0	2000	6	AX655393	AX655393 Sequence
7	38.6	4.0	249050	1	AL596165	AL596165 Listeria
8	38.6	4.0	349980	6	AX417038	AX417038 Sequence
9	38.6	4.0	349980	6	AX417041	AX417041 Sequence
10	38.6	4.0	349980	6	AX417042	AX417042 Sequence
11	38.4	4.0	2000	6	AX655393	AX655393 Sequence
12	38.4	4.0	168186	2	AC103403	AC103403 Mus muscu
13	38.4	4.0	177306	2	AC130283	AC130283 Mus muscu
14	38.4	4.0	161360	2	AC132844	AC132844 Mus muscu
15	38	3.9	214511	2	AC129256	AC129256 Rattus no
16	37.6	3.9	4526	1	AF062920	AF062920 Thermus a
17	37.4	3.8	123413	9	AL138834	AL138834 Human DNA
18	37.4	3.8	125020	9	AF429315	AF429315 Homo sapi
19	37.2	3.8	275710	2	AC135276	AC135276 Rattus no
20	36.8	3.8	125020	9	AF429315	AF429315 Homo sapi
21	36.8	3.8	349895	1	BX248359	BX248359 Corynebac
22	36.2	3.7	37106	1	AY228335	AY228335 Xanthomon
23	36.2	3.7	284984	2	AC107121	AC107121 Rattus no
24	36.2	3.7	313488	2	AC128784	AC128784 Rattus no
25	36	3.7	201627	9	AC017068	AC017068 Homo sapi
26	35.8	3.7	12396	1	AE011830	AE011830 Xanthomon
27	35.8	3.7	110000	2	LMFLCHR18_01	Continuation (2 of
28	35.8	3.7	142306	8	AP004343	AP004343 Oryza sat
29	35.8	3.7	184041	9	AC093496	AC093496 Homo sapi
30	35.8	3.7	215657	10	AL592551	AL592551 Mouse DNA
31	35.6	3.7	10211	1	AE012383	AE012383 Xanthomon
32	35.4	3.6	165400	8	OSJN00147	AL662941 Oryza sat
33	35.2	3.6	158449	2	AC115869	AC115869 Mus muscu
34	35	3.6	813	1	AB014991	AB014991 Ralstonia
35	35	3.6	1335	3	AY061932	AY061932 Drosophil
36	35	3.6	1335	3	DME302046	AJ302046 Drosophil
37	35	3.6	57303	2	AC020240	AC020240 Homo sapi
38	35	3.6	65750	9	AC092542	AC092542 Homo sapi
39	35	3.6	142101	2	AC013662	AC013662 Homo sapi
40	35	3.6	171325	9	AC079345	AC079345 Homo sapi
41	35	3.6	176610	3	AC007299	AC007299 Drosophil
42	35	3.6	270775	3	AE003624	AE003624 Drosophil
43	34.8	3.6	1590	6	AR222525	AR222525 Sequence
44	34.8	3.6	1590	6	AR432938	AR432938 Sequence
45	34.8	3.6	1590	6	AX350853	AX350853 Sequence

ALIGNMENTS

RESULT 1  
AP003594  
LOCUS AP003594 343250 bp DNA linear BCT 28-NOV-2001  
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 14/19.  
ACCESSION AF003594 BA000019  
VERSION AF003594.1 GI:17132824  
KEYWORDS  
SOURCE Nostoc sp. PCC 7120  
ORGANISM Nostoc sp. PCC 7120  
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
REFERENCE 1  
AUTHORS  
Kaneoko,T., Nakamura,Y., Molk,C.P., Kuritz,T., Sasamoto,S.,  
Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,  
Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,  
Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,

Yasuda,M. and Tabata,S.  
Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium *Anabaena* sp. strain PCC 7120  
DNA Res. 8 (5), 205-213 (2001)  
21595285  
11759840  
2 (bases 1 to 343250)  
Kaneko,T.  
Direct Submission  
Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/cyanobase//,  
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

FEATURES  
source

1. 343250  
/organism="Nostoc sp. PCC 7120"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:103690"  
/note="synonym:Anabaena sp. PCC7120"  
complement (188. .6253)  
/gene="alr3691"  
complement (188. .6253)  
/gene="alr3691"  
/note="ORF ID:alr3691"  
/codon\_start=1  
/transl\_table=11  
/product="serine/threonine kinase with two-component  
sensor domain"  
/protein\_id="BAB75390.1"  
/db\_xref="GI:17132825"

gene

CDS

/translation="MESVKMNTYVDLAIAPGYFILEBTYHSGTKTVYRAVREADQOP  
WIKLLKREYTFPSLLOPNROYAYAKNLKIPGINVLYSLPYRNSYAMWEDFGIS  
LRDYAQOOSLTETLITITIQADILHHYQRIIHKDKPANILINPETHQVKLDF  
STASLIPRETQNIISPNTLEGTLAYLSEQGRNMRGIDSDFSGLVSPELLTGE  
LPFTTDDPELHVCHIAKQDFTTIPFCQRANRFGNGESTKLPEETIPQVGE  
IVMKMAENADRYQNALGLKYLEKLAELQETGKVEYFTIGSDVCDRFITPDKLY  
GREAEITLALNFRVSLGATELMIVAGSGITGATVINEVHKPIVRQGVFIQKFD  
OFQRPNPSAFVQAPNLMGLLTSTDIQIOWRSKILEAVKNGQVIEIIEVLELI  
IGKQAPAKSLGTSIQNPNLLFKFTQVTFSAEHLPIFLDLDQWADSGSLIKQLL  
MNDTGILLIGYRDNENYNAHPMLMTVNEIGKNGTITKKDLAPNLQKNTLVLET  
LKCSBTAQSLALLVFKTQGNPFPTTQFLKALHQDNLIQNFYDILGWCQDIAQVNHQ  
ALTDVFLAKSLPRATQVQLAACITGNQDLETALVWESSPMKTAACLWKA  
LQEGILIPQSEVYFQVQEQEIDHAQETIETVYFKLHDRVQQAAYCLIPAEARIAH  
HRIGQLLKIPSAAREBHIPEIVNQLNGITLITQOSERDELAOLNNAQAARAAT  
AYQASREVADTSMTLGKTAMQROXQLTITHELAAELALLCGVEQNNQLLETVMNH  
AKCPDRVQVQVKIQAITSRNELLEAIATGSKFQELGVLSPDHPTEDEVQARBEI  
NRLIGDRLGELINIPKMTDPEKAIMQISSLIPACVMTGSLLYLLIIVLPQVLSLQ  
FNSLSFAGVYVSAYFLSTTWNMALAQGLQIAYQIASEPEAKNIRAAFTVILGGY  
FYCHTSLHQDTLPITIQEGCTAGLETGNLEFFGVYVQNFQWYNSGESLSPFASOMCT  
YHQQLDNLVNTSAMHCFYVESALILLGKSADEISLRQDAYAEKIVTEALASNDLWR  
LFOFLYRLVNLNFIADPRAKDAEAKQYLTGCMGSVAEPIEYVYDSLTALAELOS  
PAGSDSQORVRENQALIDKADHAPMNYLHKWQIVAAETHRLLEQKTEAMEFYEMAI  
KGAKENILDEALANLAAKPIYLDWGKEKVAAYVMQBAIYCYARWEAKANLDLEK  
YPQLQPLILQERQKFNFOETIATVETSSTLSLTGSTISITDLEFASILKAAQVIS  
TSLDELINLTILILENSGAKSALLIPQEDIMQIKAMTILSNFQNSSESSTQILE  
SQPLETCDIPKNIYIYVKNQQTIVDNLQTDIPGLIGEYMLQHOPOSVECTPMNO  
GHVLGILVLENLITRGVTSRLEVIRILLSQAQVSLKARLYOESOTKAQQLKOSLK  
QOKLILFNVMQROSLDNLNAJFCVVTQNIIRILLVDVRGYIQFHLVDNVEYGEFVAED  
VSPAPPALAVKQVQDHCGENYANLYKQRCITADTVQSSLELHDCRQILQIAQFVAS  
TLVPIQMBEELWGLCLICHCQDRPQWELEMQFAQVQAGMGIKQTDLLIQTOKQA  
TQLEHTLQLOQTQQLVQNEKMSALGNLVAGVAMNNPLGFTAVSLQHTQPTFADI  
MEHLKYQESLPHQDKILHAAQIDLDYILEDLPLKLEAMVACDRKNTSTSLRTF  
SRADKDYKIPFNHKGIDSTLILKHLRKADKRAPIEIVTEYGNLPHIECFPGOLNQ  
VFMNLTANAIDALEDNANGLSVAEITANNRILIRTTQVNNHVLISITDNGIGMGEEL  
KQHLFEHLFTTKGVEKGUGLGLAIARQIVVEKHGAIEVNSQLGGGTFFVIIFPTAE  
SANKY"

gene

CDS

6804. .7658  
/gene="alr3692"  
6804. .7658  
/gene="alr3692"  
/note="ORF\_ID:alr3692"

hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75391.1"  
/db\_xref="GI:17132826"  
/translation="WNFPQIGQVALALLTQFGLKVLGAIALWIVAQKLEFALKGLVR  
GLRTQNIETPLISYLLINVSIVTLIVLIVAILGFFGIETTSFALIAAGVAGVIGRWR  
GLLANFAAGAFLLIIFRPFKVGDFTIAGGITGVSGIGLFTDITTPDNVLTIVANNKI  
PADNIONFSANPRFRVDLLAQLHHDVHNQALIKARISOIPNVIENPAPDEVLITF  
NLAGPYLAVRPPYCNNDHYWQVYFDNTKAIREFGEGAGYPIPEQRQVGFSPENGTSV  
IPSSALUS"  
8079. .8939  
/gene="alr3693"  
8079. .8939  
/gene="alr3693"  
/note="ORF ID:alr3693"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75392.1"  
/db\_xref="GI:17132827"  
/translation="MNVAEISQVALALLTQFGLRVLGAIALWIIAQLLINFGKLRLR  
APRSKHVEPTLINYNVDVILKIVLIVAILGFFGIETTSFALIAAGVAGVIGRWR  
GLLANFAAGAFLLIIFRPFAVGDTIKAAGTVKVEIGLFTTITSDNVKTIIGNNKI  
FADNIONFSANPRFRVDLEAQLHHEVDHQDAIRLQKRSIQIPHVLNNPAPDEVLIF  
NLLGTVLAVRPPYCHHEHYWQVYFDNTKAIREFGEGAGYPIPEQRVAFVTCSPDGTSA  
VVORTSLGRG"  
9664. .9831  
/gene="asr3694"  
9664. .9831  
/note="asr3694"  
/note="ORF ID:asr3694"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75393.1"  
/db\_xref="GI:17132828"  
/translation="MAVDLDKRLVLILQSFHLRSWERNGLWEVSRFRRIKXHTHVSM  
VAPETELVPG"  
complement (9841. .11007)  
/gene="all3695"  
complement (9841. .11007)  
/gene="all3695"  
/note="ORF ID:all3695"  
/codon\_start=1  
/transl\_table=11  
/product="N-acyl-D-glucosamine 2-epimerase"  
/protein\_id="BAB75394.1"  
/db\_xref="GI:17132829"  
/translation="MGKNLOALQLYKNALLNDVLPFWENHSLDTEGGVFTCLDRQK  
VYDTRKFTWLQNRQVTFSLCNOLKRENMLKIASGAKFLAQHGRDARGNWYFALT  
RGGEPYLPQYNIIPSDCFAMAFSQVALASGEWAKDVAMQYNNVLRKDPKGYTK  
TYGTRPMKIDAVEMILANLHEWLEWLPQSTLENLSATVOEVMSDFDKRGLMWE  
NVADPGSHDCFEGRLLNPHGIEIAMFIMDIARRKNDKSTINQAVDVLNLTNFAWD  
KEYGQHFADYDAAGHPQQLWDDQKLMVHLESLVALAMGYRLTNGREVWYQKWH  
YSQWGFADPEYGEFGYLNRRGEVLLNLKGGKWKGCFFHVPRALYLCWQOFEALS"  
complement (11134. .12273)  
/gene="all3696"  
complement (11134. .12273)  
/gene="all3696"  
/note="ORF ID:all3696"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75395.1"  
/db\_xref="GI:17132830"  
/translation="MLPGAIALDNOGRLLVAENGTLOQVLIYKIQDPVQVGTFGHTG  
GIYQGVFGEVDLKFYGLTGVGADAQGNLYINNNFNNSGTDLRKFSQKQWELLG  
LSLVENADADPKTDGREVTFYQEQFLIDSSKPKGQQLIYKAYTLNGFKYQDSRLHIS  
PDFTFVRRLDERFLLEDLPYSNALEIYRNPSPKGNITAIAGMVFQGNIRGKSAISG  
TWPEQPTGKMIWRDNGDCGAFDKBEYDRSEDSSTYGMWVDSKGDVWFTTLGDRGI  
RHYPLQGLDNGNPIYTNISIQEKTPLIFIDLRIRIKFYPSDSDNYLSGFTAKNPPFA  
EDQAPGSEIACFDNWSKDNRLRWRIVVPKDTINKREMITASTH"

gene

CDS

gene

CDS

gene

CDS

[illegible]





REFERENCE 1 (bases 1 to 2892)  
AUTHORS Xu, G.L., Kapfer, W., Walter, J. and Trautner, T.A.  
TITLE BeuBI--an isospecific restriction and modification system of PstI:  
characterization of the BsuBI genes and enzymes  
JOURNAL Nucleic Acids Res. 20 (24), 6517-6523 (1992)  
MEDLINE 93126092  
PUBMED 1480472  
COMMENT On Feb 8, 2002 this sequence version replaced gi:263379.  
FEATURES  
    Location/Qualifiers  
        1..2892  
        /organism="Bacillus subtilis"  
        /mol\_type="genomic DNA"  
        /strain="ISB8"  
        /db\_xref="taxon:1423"  
        185..190  
        /gene="hsmBI"  
        198..21733  
        /gene="hsmBI"  
        214..220  
        /gene="hsmBI"  
        228..1733  
        /gene="hsmBI"  
        228..1733  
        /gene="hsmBI"  
        /codon\_start=1  
        /transl\_table=11  
        /product="modification methyltransferase"  
        /protein\_id="AAAI18169.1"  
        /db\_xref="GI:143054"  
        /translation="MTQILETVDKSRITVNLKNSKELQGFPTSSISIFMACLFSE  
DKLNKAVDAGVIGSLTAFALRLISENIGRADLHLEIDEMLEPYLSTLALPKD  
YIENSQIILDDFEWAAYSLLDESLANDKORFTHAILNPVKIKSKSRKLLR  
KAGETVNLVSAFVALVDLMSOGGEIVFVLPSPGCPYRPHRQHLNKLKTSIKNH  
LPFRDKAFKDDEVLQENVLSEKGVQSDKISIDSDSFSVIRYRPFKEIVOP  
NLDEKFTHINTEQLIEHPNVCSLEELNIEVSGPVVDFRVKENVLEMPGEGTV  
PLFPYHFVTSLEYEYPMOKPKPNAIIRNEKVKWLYPNHGVVYVYKFSSEKEKRIIVA  
GVLTPEVNDVDFEGLNVLHYNKSGISKEVAYGLYVLNSTPVDKYPRIENGHTQ  
VNATDLRTMKPSPSRDILISLGKWIENIENVGVQVEIDSLEELLNDRGNA"  
        1697..1701  
        /gene="hsmBI"  
        1714..2664  
        /gene="hsmBI"  
        1714..2664  
        /gene="hsmBI"  
        /codon\_start=1  
        /transl\_table=11  
        /product="restriction endonuclease"  
        /protein\_id="AAAI18170.1"  
        /db\_xref="GI:143055"  
        /translation="MTEGHSNVKEAIKILKELGPKQONERSALCLLSLMNITQDK  
TWSEAPSLIGITPMMEFCRINRYKEVAPNSRETFRTMHOQFVDAGIALYNDKPKR  
PVNSPKNVQIEATLELKYNTSEWSELLARYLNSRQTLVRYAKEROQNKIPVQI  
AEGKEIYTPGEHSELKALIEFAPRYVPGRLIYAGDTGEKMGYFDELLRLQGVV  
IDSHGKMPDVYVFPKKWLLLIESTVSHGPDVHKHBEALAKLFGNSGTAGIVVYVTFPP  
NRSIMARYLNNISWETEVWVADAPSHLIHFNHGVFLGPYE"  
ORIGIN  
    Query Match 14.08; Score 135.8; DB 1; Length 2892;  
    Best Local Similarity 47.24; Pred. No. 3.9e-28;  
    Matches 447; Conservative 0; Mismatches 497; Indels 3; Gaps 1;  
QY 19 ATCGACGGAACGGTAGCGACATCGATCTGCGCGGCGCTGCTTAAAGCGTTTGGGTTT 78  
Db 1717 ACCGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1776  
QY 79 GACCGCAACGATATACGTCGCTAGCGCTGTGACATGCTCGCGCTTGGCGTTTGAAG 138  
Db 1777 CCTAAGGGGAGCAAAATGAACGTTACGCTCTATGTTTATCTTAAATGAATAACT 1836  
QY 139 CCGGAGATCGTGGGTTGACTGACCACTCCAGCCCTTGGCGTTTGAAGATCATGGAC 198  
Db 1837 CAAGATAAGACATGGTCAGAGAGCTGAAAGTCTCTTATTGGTATACCAACATGATGAA 1896

QY 199 TGGTCCGCGAGCATTGGCCCAAGCCGTAGCCACCGGAGCTCGAGAGATTTCGGTAAG 258  
Db 1897 TTTTCCGCAATTATATGAAAAAGATATATGCAAAATAGTCGAGAACTTTTAGAAGG 1956  
QY 259 AAGACGCTTCGCGCAGTGGGTTGATAATATGCGCTTCGCGGTACTTAATTCGCGCAATTTAAAC 318  
Db 1957 TTCACTATGATCATGTTTGTAGATGCTGGAATTCGTTGTATATACCTGTATACCCGCAACCGACA 2016  
QY 319 ATCCCGCAGAACTCGAGCTCAACGAGTACTGCTGTGCTGACGAGCAATTCAGCGCGTA 378  
Db 2017 AGACCTGTTATATAGTCCGAGGAGTATATCAAAATAGAGCGGAACTTTAGAACTTATT 2076  
QY 379 AGGCGATATGGAACGGAAGCTTCGAGAAATCTCTGTAGTCTTCTTGTATGAGAGCATCG 438  
Db 2077 AAGTGCTATATACAGAAGAAATGAGTGAATTAATAGCCCGATCTTATCTAATAGACAA 2136  
QY 439 AAGCGGTTTAAAGCGCGAGGAAAGCTCTCCAGGCTCGATGATCTCTGTGATCTCCCT 498  
Db 2137 ACATTGCTAGAAAGGTATGCTTAAAGAGAGACAGCAGATAAAATACCTGTACAATAGCT 2196  
QY 499 GGTGGCAGGAATTTCTGCTCTCGCCCTGCGCGGCGAGAAATCCATTGCTGAAGAAAGTGGT 558  
Db 2197 GAAGTTAAAGAGATTTATATAACTCCAGAGAGACACAGCGAGTTAATTAAGGCAATTATA 2256  
QY 559 GAAGAGTTTGTGCGCGATTTGCACCTCGCTCGAGCGGTCTACCTCGGGGATACCTCGT 618  
Db 2257 GAAGATTTGCACCCCTTATGTAACAGGAGGAGATTAATTTATGCTGGTGATCTGGT 2316  
QY 619 GGAAGCAATTCCTTATTCGAACGAGAGATCTTTGAAGAGGTGCTCGGCGCTGACCTTCGAC 678  
Db 2317 GAAAGATGGTATTTTGTATGAAGATTAATTAAGACAGTTA---GGGTAGTTATAGAC 2373  
QY 679 CCCATGCTGAAATCGCGGACCTTATTCCTCCATGACGAAGTTCTGGTGGTGGCTTTTCCCT 738  
Db 2374 TCGCATGGAATAATGCTGATGATGATTAATTTATTTCCAGAGAAAAAGTGGCTCCTACTC 2433  
QY 739 ATGAGAGCCGTGAAAGTAAGTAAAGTCCGTTTGTAGAGAGCGGATCCAGCCTCGAAGAG 798  
Db 2434 ATGATCGGTAACTAGTCACGGTCCAGTAGACCATTAAGCGTCAAGAAATTAAGCCAAA 2493  
QY 799 CTATTCGTTACACCTTCAGCGGCTCTAATTTTGTAAACTGCTTTGAAATTCGTGAGTCG 858  
Db 2494 TTAATTAATGTTCAACTGCGGGAATAGTTTATGTGACTCAATCCCTAATCGTTCAATTA 2553  
QY 859 ATGCGTCAGTGGCTCCCTGAGCTGGCTTGGAAACTGAGCGGTGGGTAGCGGAAGATCCA 918  
Db 2554 ATGCGAAGATACCTTAATAACATTTCTTGGGAGACGGAAGTGTGGTAGCTGATGCACCG 2613  
QY 919 GACCATCTGATTACCTTAACGGGTCTAGATTTCTTGGCGCGTACGA 965  
Db 2614 TCACATCAATTCATTTTAATGGCGTTCGTTTCTTAGGCCCATACGA 2660  
RESULT 4  
PROJRM  
LOCUS  
DEFINITION P.stuartii PstI restriction and modification genes, complete.  
ACCESSION K02081  
VERSION K02081.1 GI:150922  
KEYWORDS endonuclease; methylase; modification enzyme; restriction endonuclease.  
SOURCE  
ORGANISM  
    Providencia stuartii  
    Providencia stuartii  
    Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
    Enterobacteriaceae; Providencia.  
REFERENCE 1 (bases 1 to 3889)  
AUTHORS Walder, R.Y., Walder, J.A. and Donelson, J.E.  
TITLE The organization and complete nucleotide sequence of the PstI  
restriction-modification system  
JOURNAL J. Biol. Chem. 259 (12), 8015-8026 (1984)  
MEDLINE 84239756  
PUBMED 6330092  
COMMENT Original source text: P.stuartii DNA, clone pPst201.

The two genes for the PstI restriction-modification system are encoded on opposite strands and are transcribed from separate promoters. The restriction endonuclease contains 326 amino acids. The modification enzyme contains 507 amino acids. The close proximity of the transcription initiation points of the two genes (70 bp) suggests that the promoters overlap. The -10 regions for the two genes are found at bp 2307-2302 (MME) and 2352-2357 (PstI). Bases 2267-2340 are protected from DNaseI digestion. The PstI restriction-modification system is a type II system. The two enzymes both recognize the hexanucleotide sequence 'ctgcag'. The modification enzyme protects this sequence from cleavage by PstI by methylation of its adenosine residue. PstI cleaves the DNA between the adenosine and guanine residues, leaving 'sticky ends' of 4 bp.

```
FEATURES
  source
    1..3889
      /organism="Providencia stuartii"
      /mol_type="genomic DNA"
      /db_xref="taxon:588"
      /complement(<742..2292)
      /product="MME mRNA"
      /complement(742..2265)
      /note="methylase modification enzyme (MME)"
      /codon_start=1
      /transl_table=11
      /protein_id="AAA25672.1"
      /db_xref="GI:455323"
      /translation="MTEAATLPISLILVDSIRAA NSTLDLTRSKLQFMSSSAV
      SELMANLPSYGEHEILDAGVGSILTAAFVONATLNGAKSISCTCYEISRYMYNL
      IQVLDLCKIRAMEFVNWQKILIESDFOASVEQLLIENYSKYNKAILNIPYLKIA
      KGRALDQKQVIEASNLVAFVLAIKQLKSGELIENYSPNGFPNDFRKLAA
      DECSLNKHVNSRFSKAFKADNLQENIYIHLTKGTQRKVTVYSTCANDINTIF
      EYFPEIIVKNSPLIFIHVTNEQREILANKAGGLPCSLDGIQVSTGVNDFFRE
      NLSMEYINSVPLIYPPHOLQRCISVWPTFKAKPNALVLEATNLMPNGIYVLTTR
      LTAKEKRIVASIYPIPLANDVTGDFONKYNFHANGKPLDISLAKGLVPLNSTLI
      DKYFQMNGHIVADIDRALRYPTREQLIEDIANQVDFGEFEQTKIDEILNQSILQM"
      /size=33376
      /product="PstI mRNA"
      /size=3376
      /note="PstI restriction endonuclease (PstI)"
      /codon_start=1
      /transl_table=11
      /protein_id="AAA25673.1"
      /db_xref="GI:150924"
      /translation="MKELKKEAKETILKALGPLPQOYNDRSGVLLALANIKPESWK
      EKAPLLPTVIMEIRTEYQKYPNSRETIROTLHOFQEARIVDRNRDLPSSRATN
      SKDNYSLNQVLIIDILHNPYNGWKELIQOFLTHVPSLQELYERALARDRIPKLLDG
      TQISLPGHNLHADIHVEFCRFVGMGKILYIGDTASSRNEGKMLVLDSEYLKK
      LGVFPMSHDKLFDVVYDEKRWLFLEAVTSHGPISPKRWLELEALSSCTVGVYIV
      TAPPTRTFRKNAANIAMEYVIADNDPMVHFNGDRFLGPHDKKPELS"
      /note="1 bp upstream of HindIII site."
  ORIGIN      1 bp upstream of HindIII site.
    Query Match      6.0%; Score 58.8; DB 1; Length 3889;
    Best Local Similarity 50.3%; Pred. No. 1.4e-05;
    Matches 144; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
  QY      680  CCCATGGTCGAATGCCGACCTTATTCCTCATGACGAAGTTCGTGGTGGCTTTCCCTTA 739
    Db      3069  CACACGACAAITGCCAGATGTTGTTGTACATGAAAGCGTAATGTTATTTTAA 3128
  QY      740  TGGAGCCGTGAAAAGTAAGTTCGTTTCATGAGGAGCGGCATCCACCTGCAAGAGC 799
    Db      3129  TAGAAGCGGTTACATCCCATGGCCCAATTAGTCCTAAACGTTGGCTAGAACTTGAAGCTG 3188
  QY      800  TATTCGTTACACCTTCAGCGGGTCTAATTTTCTAACTGCTTTGAAAATCGTGAGTCGA 859
    Db      3189  CCTTGTCACTCACTGACGTCGGTAAAGTGTAGTTACTGCTTCCCAACACGTACAGAGT 3248
  QY      860  TCGCTCAGTGGCTCCCTGAGCTGGCTTGGGAACTGAGCGTGGGTAGCGGAAGATCCAG 919
    Db      3249  TCAGAAAAAATGCGGCAAAATTGTCATGGGAAACCGAGGTTTGATTTGCTGATAATCCAG 3308
  QY      920  ACCATCTGATTACCTTAACGGTCTAGATTTCTTGGCGGTACGA 965
```

```
Db      3309  ATCATATGGTCATTTCAATTCAGCGACCGTTTCTTAGGCCACATGA 3354
  RESULT 5
  AE016775/c
  LOCUS      300511 bp      DNA      linear      BCT 12-DEC-2002
  DEFINITION  Pseudomonas putida KT2440 section 2 of 21 of the complete genome.
  ACCESSION  AE016775 AE015451
  VERSION    AE016775.1 GI:26557018
  KEYWORDS
  SOURCE      Pseudomonas putida KT2440
  ORGANISM    Pseudomonas putida KT2440
              Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
              Pseudomonadaceae; Pseudomonas.
  REFERENCE
    1 (bases 1 to 300511)
      Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
      Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
      Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
      Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Uterback,T.,
      Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
      Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
      Duesterhoft,A., Tummler,B. and Fraser,C.
      Complete genome sequence and comparative analysis of the
      metabolically versatile Pseudomonas putida KT2440
      Environ. Microbiol. 4 (12), 799-808 (2002)
    2 (bases 1 to 300511)
      Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
      Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
      Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
      Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Uterback,T.,
      Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
      Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
      Duesterhoft,A., Tummler,B. and Fraser,C.
      Direct Submission
      Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
      Medical Center Dr, Rockville, MD 20850, USA
  FEATURES
    1..300511
      /organism="Pseudomonas putida KT2440"
      /mol_type="genomic DNA"
      /strain="KT2440"
      /db_xref="taxon:160488"
      /complement(80..985)
      /gene="rimK"
      /locus_tag="PP0248"
      /complement(80..985)
      /gene="rimK"
      /locus_tag="PP0248"
      /note="similar to GB:X57130, and PID:296288; identified by
      sequence similarity; putative"
      /codon_start=1
      /transl_table=11
      /product="ribosomal protein S6 modification protein"
      /protein_id="AAN65880.1"
      /db_xref="GI:24981607"
      /translation="MKIIVLSRNPRILYSTRRLVEAGTQRGHEWVITLIRAYMNIASH
      KPIIHVRGKPLEGFDAVPRIGASVTFGCAVLRQPMGVVPLNESVIAISRDKLR
      SIQLLSRGIGIPITGFAHSPDDIPDLIVNGVAGPLVIVLSTGQIGVVLCTPOAA
      ESVIEAFMGLKQNIWQVEYIKAGGADIRCFVVGKVIASMKRQAKPGFRGNLHRGG
      VASLKITPEERTAIRAAKVMGLSVGVDDILRSNHPGLVMEVNSSPGLEIEVITGK
      NVAGMIIEHLKNGGNPQTRTKGK"
      /complement(982..1602)
      /locus_tag="PP0249"
      /complement(982..1602)
      /locus_tag="PP0249"
      /note="identified by Glimmer2; putative"
      /codon_start=1
      /transl_table=11
      /product="conserved hypothetical protein"
      /protein_id="AAN65881.1"
      /db_xref="GI:24981608"
```

```

/translacion="MSSSSAFCAIPLPRILFIILYIHLIRMQKPAVLKVHLLIRRV
RFVTKTDFHIVIGREWVALPDVAGLRAKIDITGASTSSLHATEVEPFRDGGOPWVR
TAHLGSLVQVLRHRCRAPVTMTKTIKSSNGHACTRVYIIFSLAGGVWEVETLAC
RKNRYRLILGSKALIHQQLVNNGLKYVQKQPAFPATLSPVTGAA"
1574. .1975
/gene="hs1R"
/locus tag="PP0250"
1574. .1975
/gene="hs1R"
/locus tag="PP0250"
/notes="identified by match to PFAM protein family HMM
PF01479"
/codon start=1
/transl table=11
/product="heat shock protein 15"
/protein id="AAN65882.1"
/db_xref="GI:24981609"
/translacion="MAQAKDDDDKVRLLDKMLWAARFYKTRALAKAAIESGKVHCRGR
KPKGKPRVGDGFLVITGDFERTVVVKALSVYRRGAPEAQAALYEETESVRRREQAAE
MRKAGAMGVTTDGRPTKKORRQIHLHGSYD"
complement (2004. .2798)
/locus tag="PP0251"
complement (2004. .2798)
/locus tag="PP0251"
/notes="identified by match to PFAM protein family HMM
PF01569"
/codon start=1
/transl table=11
/product="conserved hypothetical protein"
/protein id="AAN65883.1"
/db_xref="GI:24981610"
/translacion="MDNRPLFOARWSWGPIAALLPIALLCFWLPFGOILCLTFDE
WLPRLNAPLADSLTMYIWTIGSLRPEDIVGVLILAVLRIGDWPVKAQVRAFPFG
FLVTLLLVIVRALFISKWDAAGQHQKSPMIFDDVHLSDYFYNLRAELKDRSK
SFDGASVLLIHALFMVSRGLAQIWMVGLAALPRLVAGAHWGQDDYTGGLL
MAVLAWGVSYYTTPLAAGSEALMRWTAPLKLSTLPLVGRMAVR"
2919. .3818
/locus tag="PP0252"
2919. .3818
/locus tag="PP0252"
/notes="identified by match to PFAM protein family HMM
PF01430"
/codon start=1
/transl table=11
/product="chaperonin, 33 kDa"
/protein id="AAN65884.1"
/db_xref="GI:24981611"
/translacion="MSDLPTDTQRFIFDERIVRGSEWVSLDSDS YAAVLARHEYQPQV
QVLIGELMAATALIVGAMKFDGLILQARGAGPIPLIMVCSEREIRGMARYEADQI
PAGATLSQLPDGHLLTIDPVKGQRYGTVLDGANSLSCTFTNYFVQSQOLNTRFLW
NAQGGKARGILLQOLPRDOPDDEERDSQHVVAKTLKPKREWTEGNETLLHRLYH
EDAVRLFDIQLRPNCSRSRSGNALVSLGEHDAKALVDECGTVEDICQFCNERIF
FDASDVQLFAGGGTDVASETRH"
3998. .5537
/gene="pckA"
/locus tag="PP0253"
/notes="This region contains an authentic frame shift and
is not the result of a sequencing artifact;
phosphoenolpyruvate carboxykinase, authentic frameshift;
similar to GB:U04815, GB:U04818, GB:U04824, GB:U04816,
GB:U04817, SP:P21127, PID:189481, PID:507158, PID:507160,
PID:507162, PID:507164, PID:507166, PID:507168,
PID:507427, and PID:507429; identified by sequence
similarity; putative"
5686. .6126
/locus tag="PP0254"
5686. .6126
/locus tag="PP0254"
/notes="identified by match to PFAM protein family HMM
PF03061"
/codon start=1
/transl table=11
/product="conserved hypothetical protein TIGR00051"

```

```

/protein id="AAN65885.1"
/db_xref="GI:24981612"
/translacion="WTEAPORTAFSHFPIILTRQDNDNLNGHIIAGATVGHGFETAIQA
FLVQAEILDRHGLAFLVSSAADFYALPGFDDLLEVLGVTLRAGSTVYRLALYR
PGSDACAAGTVQVQVFIERASGREVALPEALQVILSGVLERQA"
complement (6183. .6665)
/locus tag="PP0255"
complement (6183. .6665)
/locus tag="PP0255"
/notes="identified by Glimmer2; putative"
/codon start=1
/transl table=11
/product="conserved hypothetical protein"
/protein id="AAN65886.1"
/db_xref="GI:24981613"
/translacion="MKYKISILLISFLLSGVMAGTTTGAIGGALGGVILGWSVWNS
IGSTGAIGALGAGGAGLADKQKRGGAIIAGGALGAGNNGVSGVGGTTSYIG
AAAGGGAGGALGNYLGKADKDDYRDGRRYRGYDDDRHWDGRGHYGRKHKHRWD
"
complement (6862. .9207)
/locus tag="PP0256"
complement (6862. .9207)
/locus tag="PP0256"
/notes="similar to SP:P07658; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="molybdopterin oxidoreductase, alpha subunit"
/protein id="AAN65887.1"
/db_xref="GI:24981614"
/translacion="MTSQQLPONTPASPPRYKPYHGAGGALRSVAKAWGSDNA
LKNIRALLKTNQGGEDPCSCMGWSDPSKRVFCENGAKAVWEATKRRVDAAPFAR
YVTSLLQSDIWMLEYOGRTEPEVYDAPSDRVLPIFISWDAFALIAELANLPDQA
EFTYSRGAAYALYQLFVRYATNNPDCSNMCHASGVALGQSVGKGVTFDD
FEHADAI FVWGQPNTHPLDPLRDVAKRGQVVCINFLKESGLERFQHPQPLEM
LITNSDRNTAFTRPALGQDMALRGKAFVLOWERAQNGEPAVDFHAFIABHGH
VDEYLVADATPWSHIOAQSLTLADIELAARYCQKRVIMCWAMGITOHRHSVPTI
QEIVNOMLRGNIGVPGAGLCIPVGHNSVQDRTMGINERPPALLDAIEKRFQFPVP
RRGNHIVEALHMLGRKAVFGLGNGFAQATPDTERTAQRNCLBLVHISTKLMR
SHLVHQKALILFCLGRTDIDLDQDGPVAVTDEDSFMSHASNGQLKPLSTQMRSEPA
VIAGIAAATKQFQVDMHVLVADYRIDLIGDTIPGFSGNRQRLNPPGSGFLGNSAA
SRWATSTGRANFKANLLPTLLDERVRASGLPDLIMQSMRSHDQVNTTYIGLDDRY
RGVQGRVLFANEADIIIRLGFGQKQKVDIVSLWGDEHVRVQGFLLADIPAGQAA
AYVPEVNPVPLESIGVSGSTPSKFIKLERAREDRIL"
complement (9213. .10043)
/locus tag="PP0257"
complement (9213. .10043)
/locus tag="PP0257"
/notes="similar to GB:X60787, GB:S41456, SP:Q01167,
PID:1388150, and PID:33854; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="formate dehydrogenase accessory protein FdhD"
/protein id="AAN65888.1"
/db_xref="GI:24981615"
/translacion="MNSKPEVCAASSPLSNPAASTNYVYVQLSDAKRDSPLAEVAL
AIYVGNQAMLVSPDLEDFAVGFSVGIVGTAETIYDLKLSGSGSALYADLEIS
SRAFWNLKNRRQRLAGTSGCGLCGVEALEQALPDALVLPAGLPPAQLVGLRERIDA
"
Query Match 4.7%; Score 45.4; DB 1; Length 300511;
Best Local Similarity 54.5%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 76;

QY 645 GATCTTTGAAGAGTGTCTCGGCTGACTTTCGACCCCATCGATGCGGACCTTAT 704
Db 31147 GCTGTTCCGGCGACTGCATCGCGTTCGACCTGTGTCGTTTTCACGACGAAT 31088

QY 705 TCTCCATGACGAAGTTCGTGGGTGGCTTTTCCTTATGAGAGCCGTGAAAAGTAAAGTCC 764
Db 31087 GCAACACGACGACCATGTCTGCAAAATTTGGCCACATACACGCATCGAATGCGAGGCTGC 31028
QY 765 GTTTGATGAGGAGCGGCATCGACCTGCGNAGAGCTATTCTGTTACAC 811

```



KEGHDLQIKLILEVKKGSBLDDVKVEVDNQIKNHLRLKVNKTGKTDYNIASLPFGY  
IERKVPVANNQBTYSMPDIDIEPLQSVTLTNEAESCTTFTQGVKEYQOLDOHLALT  
LLATLTQGLDPLAYPRASGDPTTKGHVLIETEGRAQLLGKHEFTLAIYLGADPE  
HKTAERTKAFQANVSQVORPPHFHLRLDNKLOKTERKLGAKQTLGMLNPKXYLVS  
ACHPSYTTADKYIIRLENPTNTPOKLPLDVPSPFYNVAIEEVVANDNNTTIPAYGAVTL  
RLDL"  
3927. .3954  
/gene="lin0449"  
complement(3929. .4816)  
/gene="lin0450"  
complement(3929. .3956)  
/gene="lin0450"  
complement(3941. .4816)  
/note="similar to transcriptional regulator (LysR family)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95682.1"  
/db\_xref="GI:16412891"  
/db\_xref="GOA:Q92EK9"  
/db\_xref="SPTREMBL:Q92EK9"  
/translation="MELRVLYNLYLTVAEREKTIISPAEVLHLSQPTLSKQLKELEELG  
VTIFRGNRSITTEDGLYLANRKEILSLVELTSLNLLQNETISKIMIGAAETRGF  
EFTGVLHEUREHPDITFQHSGNADDDVLEKIEHGLLDFGLVINPAEKLOYELPLT  
IEDTWGLVNTKRPDLAEKVVVTPSDIDANPIVWSQSFIDNQLASWLGHFQELNVITG  
RYLLYNASLLAKENIASILCIDGIINTENTNLKFPFSPPLTAGISIVWKKNTQFS  
ASKEFLRLIOKTCSK"  
4934. .5503  
/gene="lin0451"  
4934. .5503  
/gene="lin0451"  
/note="similar to acetyltransferase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95683.1"  
/db\_xref="GI:16412892"  
/db\_xref="SPTREMBL:Q92EK8"  
/translation="MIBQDLQKRLVQEIILQSGDLLDEIHEVKONNEQLIIELNKKYH  
TKREVTKYLSLDITKQKVPDPSVDVSLPSPDFGKHITFGKNLFLNLNVTFDVLDGLITLE  
DNVLIGPARLVTYVHLVSPKRGRLVAFICVKNAMIGANATILSGVIGENAIVA  
ADATVKDVPANVIVAGSPAKQIRKIIIE"  
join(5503. .5508,5517. .6263,6272. .6302)  
5503. .5508  
/gene="lin0452"  
5517. .6263  
/gene="lin0452"  
/note="similar to oxidoreductase"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95684.1"  
/db\_xref="GI:16412893"  
/db\_xref="GOA:Q92EK7"  
/db\_xref="SWISS-PROT:Q92EK7"  
/translation="MTIKNKVIIITGASGIGTEATALLAEKAGKVLAAARRVEKLEK  
IVOTIKASGEALFAKTDVTKREDNKKLVELATERYGVKVDALFINAGIMFNSPLSALK  
EDWEOMIDINIKGLNGIAVLPSFLAQSKHILATSSVAGLKAYPGGAVYGATKWA  
VRDLMEVRMESAQEGNIRATIPAAINTELLETITIDKTEQGMNLNLYQYGVTPD  
RIASIVAYAIDQPDIDNVNEFTVGPPTQPW"  
6272. .6302  
/gene="lin0452"  
6362. .6884  
/gene="lin0453"  
6362. .6367  
/gene="lin0453"  
6375. .6884  
/gene="lin0453"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95685.1"  
/db\_xref="GI:16412894"  
/db\_xref="SPTREMBL:Q92EK6"  
/translation="MNRKQLEQLDKCLYEYVDEQETNDFRVGYIRAFIDFSFIILEK

terminator  
gene  
terminator  
CDS

gene  
CDS

terminator  
gene  
RBS  
CDS

YHTRVGFDPGALVANEAFKVOQDIOYLVKPYQNLRRPQEQIPELSEDDGDTLLNPLEWT  
KKNKLVEIMTVWDELLIETIETIENLVMNTVAPEEFQSDGKCFTECDNIVDICID  
TABLKRTT"  
7123. .13641  
/gene="lin0454"  
7123. .7128  
/gene="lin0454"  
7138. .13641  
/gene="lin0454"  
/note="similar to cell wall-associated protein precursor  
wapa (B. subtilis)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC95686.1"  
/db\_xref="GI:16412895"  
/db\_xref="GOA:Q92EK5"  
/db\_xref="SPTREMBL:Q92EK5"  
/translation="MKIHAKRNVLSLIAFMFLFSLPGYLAEDSPVGVADPDKK  
AGEQAPTEVEKERTENEVFDNHDGSPFKQIFADPINMEVDGMKRIADVKEADSD  
MIVPKOTPLEGLEKMEGDGAYOKLTKAGAEVTFRLKGARTGENEOAVTQPAITYKEN  
EVTYENVPPTDLRHLTPQSVKEDLVLRPNQVDTVYQIETKLDELAENGDLVLEK  
NKSDETMYTILPKPMTDSNVGTETGEAAQSENVSFKVQLTKTYELQLKVDREWLND  
DAREYPYIDPSVRLDKINYANINSAPKETETNIGSKLWDSQNAVITLUGKWDNSTGN  
NAAPLKMOTSLNKATISAKTLKVNIWHSPTVNDLWYENAKMSPPQVWNTVTP  
PVTRLGSVNVGRGEWANLDVTKTQAWASGSRVNGFRLGTNINKNWKVQVASENNK  
NYPLEVNYTYAQPEKPTVKTSSNGVTGTGMDLSGKAVGATSYNIVISNGVKYE  
FNTKTATWTSTGKKI FPTNDRIAKGERFEHDKGTDFALDPRQAYENAFQAGSTF  
GLNLRILYLFVQAVPFGSGSPNSDLVFAVMPLEKQAPVAKAYSNLAHKTGYVELN  
WEKSPWADGGYKLVLFVQAEYEDVGAETKTQTKNGKIWFPTKEIABGKALHHDGKG  
AELAKOPSPVYTSNGGYKERTNYFVRIAYQKAGNNNTSVQSEPAFPTPIEALNKQL  
GMVDYMTSPVRGGEVNAINGFLFHTDFNLGEGPSINVNRTFNSQDDATGIFGKG  
WTSLEBKLEVEENGNIWVYESDKIHFRTKGDKEYEAPPGIYSEITKNADGVLKIEE  
DKSETPLVDGRLEKSEKDTGNELTYEYTGKLTSLDASGRVTTLTYEGELVKELVG  
PEDRKISYTYNDKQELISSSTARGKLYRYGTGTLTSLSYDPKHTKEKPYETPAYEE  
EKLTETIDPVGKKTIYSYDRAEQOITLNEKKKRTIYSYNDAGNPKKEIWDADGLKT  
TITYTESNLLVKEVNPKGQETIYADADGNITKATDAYTESYTYNDNNDVTSSTDTE  
GRKTYTYDGADAVSETLATESQVSSVYQDAYGNPIRSGSELSNGSLQNSGFEGE  
AGVSNWTLIOSDAKGSMTFNTOSAPGALGSGSVKLTSEANSYKGVSSVYQVDE  
PETYTYFSAMIKTSGMTNADALLIGRLQDANAKDVTDAGVQSNRATS.IKKGDMVKR  
QLTFKSKNTRQVLLVLDNEQAPHPKGGIITWDNVQFEKGSVASSYNPVPVNSFENH  
NGTLPTGMKRTGNALTQAKVDNQSNGSDSAVYFERKATSEATHTIVQDVPVNNKSA  
KALTIKALSKESEDAKANGSVATMSNDYSVMGTVYYQDGTSSVQGGPLGTNDNMRSA  
VVVKPTKPVKMIKVYTFMRNGLTKAWDFDVRVEGELTKNEYDASGNVYATSYDSE  
GRKISFTYDIYGNKTSETDEKGNKTLTYDADNALIDTKLANGTSVAYKYDDNGNTTE  
XNVTASGKTQKNIYEYVDVNKITAFTDALNRITIKYEYDAAGNETKALMPNGRVTESY

Query Match 4.0%; Score 38.6; DB 1; Length 249050;  
Best Local Similarity 61.4%; Pred. No. 16;  
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 453 GCGAGCGGAGCTCTCCAGGCTGGATGATCTCTGTGATCTCCCTGGTGGCGAGGAATT 512  
Db 147936 GCAGCCCAAAATTGGCAGCGCTGATGAAATTTATCAAGTTGGCGTGGCGACGGAATC 147877  
Qy 513 TCTGCTCTCCCTCCCGGCGAGAAATCCATTGCTGAAGAAGA 553  
Db 147876 GCTGCCCTAGCTACGCTACGATCCATCTCTTAAGTAGA 147836

RESULT 8  
AX417038/c  
LOCUS AX417038 349980 bp DNA linear PAT 02-SEP-2002  
DEFINITION Sequence 4029 from Patent WO0228891.  
ACCESSION AX417038  
VERSION AX417038.1 GI:21449648  
KEYWORDS  
SOURCE Listeria innocua  
ORGANISM Listeria innocua  
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
REFERENCE 1  
AUTHORS Kunst, F. and Glaser, P.  
TITLE Listeria innocua, genome and applications  
Patent: WO 0228891-A 4029 11-Apr-2002;  
JOURNAL









[illegible]

```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-136E18"
/clone_lib="RPCI-23 Female Mouse BAC"
1..65842
/notes="assembly_fragment
vector_end:SP6
vector_side:left"
65943..68847
/notes="assembly_fragment"
68948..73881
/notes="assembly_fragment"
73982..81087
/notes="assembly_fragment"
81188..108661
/notes="assembly_fragment"
108762..140549
/notes="assembly_fragment"
140650..177306
/notes="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match      4.0%; Score 38.4; DB 2; Length 177306;
Best Local Similarity 57.5%; Pred. No. 18;
Matches 69; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 749 TGAAGTAAAGTCGCTTTGATGAGGAGCGGATCGCAGCTGCAGACCTATTCTGTTA 808
Db 15161 TTAATATGGAGGATCATATGAAGAGAGAGATTATGCGCAGCATCCAGACCTATTTTGT 15102

QY 809 CACCTTCAGCGGCTCTAAATTTTGTAACTGCTTTGAAAATCGTGAGTCGTCAGT 868
Db 15101 TTCCCTTGGATGATTTATTTTGAATCTGTTTTTTAACTTGAATGATGTCAGT 15042

RESULT 14
AC132844/c
LOCUS      181360 bp      DNA      linear      HTG 16-FEB-2003
DEFINITION Mus musculus clone RP24-481P13, WORKING DRAFT SEQUENCE, 3 unordered
            pieces.
ACCESSION  AC132844
VERSION     AC132844.3 GI:28394960
KEYWORDS    HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 181360)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Garmy,L., Gordon,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Linderblad-Toh,K.,
            Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
            O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
            Direct Submission

TITLE

```

## JOURNAL

REFERENCE  
AUTHORS

Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 181360)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Linderblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 16, 2003 this sequence version replaced gi:25956358.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27224

Center clone name: 481.P.13

----- Summary Statistics

Sequencing vector: P1asmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 180535 bases at least Q40

Consensus quality: 180783 bases at least Q30

Consensus quality: 180973 bases at least Q20

Insert size: 181160; sum-of-contigs

Quality coverage: 11.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved

\* 1 103657: contig of 103657 bp in length

\* 103658 103757: gap of 100 bp

\* 103758 110625: contig of 6868 bp in length

\* 110626 110725: gap of 100 bp

\* 110726 181360: contig of 70635 bp in length.

Location/Qualifiers

1..181360

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP24-481P13"

/clone\_lib="RPCI-24 Male Mouse BAC"

1..103657

/note="assembly\_fragment

vector\_end:SP6

vector\_side:left"

103758..110625

FEATURES  
source

## misc\_feature

## TITLE

## misc\_feature

misc\_feature  
110726..181360  
/note="assembly\_fragment"  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right

## ORIGIN

Query Match 4.0%; Score 38.4; DB 2; Length 181360;  
Best Local Similarity 57.5%; Pred. No. 18;  
Matches 69; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
Qy 749 TGAAGAAGTAAAGGTCGCTTGATGAGGAGCGCATCGCAGCGCTGCAAGAGCTATTTCGTTA 808  
Db 15169 TTAATATGAGGATCATATGAGAGGAGATTCATCGCAGCATCAAGACCTATTTTGT 15110  
Qy 809 CACCTTACGCGGTCCTAATTTTGTAACTGCTTTGAAATCGTGAGTCGATCGCTCAGT 868  
Db 15109 TTGCCTTGTATGATTTTATTTTGAATCTGTTTTTTAACTGATGATGTCGTCAGT 15050

## RESULT 15

AC129256  
LOCUS Rattus norvegicus clone CH230-384015, WORKING DRAFT SEQUENCE.  
DEFINITION Rattus norvegicus clone CH230-384015, WORKING DRAFT SEQUENCE.  
ACCESSION AC129256  
VERSION AC129256.4 GI:25072556  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus

## REFERENCE

1 (bases 1 to 214511)  
Munzy,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleaveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,  
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,  
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorenshuwa,L., Loulseged,H., Lozada,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,  
Nwaokemele,O., Okunolu,G., Olarnpusagoon,A., Pal,S., Parks,K.,  
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,  
Plopper,F., Poindexter,A., Popovic,D., Prismus,E., Pu,L.-L.,  
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,  
Weinstock, G. and Gibbs, R.A.  
Direct Submission  
Unpublished  
2 (bases 1 to 214511)  
Worley, K.C.  
Direct Submission  
Submitted (28-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 214511)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23815005.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GTGP  
Center clone name: CH230-384015  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 201412 bases at least Q40  
Consensus quality: 203652 bases at least Q30  
Consensus quality: 205020 bases at least Q20  
Estimated insert size: 208107; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces  
is believed to be correct as given, however the sizes  
of the gaps between them are based on estimates that have  
been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 214511: contig of 214511 bp in length.

FEATURES  
source

Location/Qualifiers  
1..214511  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-384015"  
misc\_feature 1..1753  
/note="wgs\_end\_extension  
clone\_end:Sp6"

```

misc_feature      3142..4830
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      6595..7733
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      8697..9302
                  /note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ281669"
misc_feature      complement(207915..208580)
                  /note="clone_boundary
clone_end:T7
site:
end_sequence:BZ281668"
misc_feature      211140..214511
                  /note="wgs_end_extension
clone_end:T7"

ORIGIN
Query Match      3.9%; Score 38; DB 2; Length 214511;
Best Local Similarity 47.8%; Pred.No. 23;
Matches 110; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY      513 TCTGCTCTCGCCTGCCGGCAGAAATCCATTGCTGAAGAAGATGGTCGAAGAGTTTGTGCC 572
Db      136957 TCTGGAGTAGAGAGAAGGTGTGCAGAAAGTGGGGGAGAGCTGGGCTTCAGCTAGGGCTG 137016
QY      573 GCGATTTGCACCTCGCTCGACGGTGCTTACCTCGGGGATACCTCGTGGAAAGCATTCCT 632
Db      137017 GAGTGTGGGAGTTGCTCTGGGCTGGCATCCTACTGAGATACCTGGAACACTATTCCAGT 137076
QY      633 ATTGGAACGAGAGATCTTTGAGAGGTGCTCGGCTGACTTTCACCCCATGGTCGAAT 692
Db      137077 CTCAGAAGGAAGTTTCCAGGAGATCTGGATGTCCTGTGTTGGATGCTACTAGACTGAG 137136
QY      693 GCCGGACCTTATCTCCATGACGAAGTTCGTGGGTGGCTTTTCCTTATGG 742
Db      137137 GCCAGAGATGCTCTGTCTTCCTATCGAGTACCTGATGCCCATCCTCCAGG 137186

```

Search completed: October 2, 2004, 03:13:40  
 Job time : 3966.41 secs

P Lank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 19:21:23 ; Search time 416.286 Seconds  
(without alignments)  
9919.257 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972

Sequence: 1 gtgacagcagctgacgcat.....ttgggcgctagcaacgtag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.6	4.0	2000	7 ADA71938	Ada71938 Rice gene
C 2	38.6	4.0	110000	6 ABQ69245_05	Continuation (6 of
C 3	38.6	4.0	110000	6 ABQ69245_06	Continuation (7 of
C 4	38.6	4.0	110000	6 ABQ67197_05	Continuation (6 of
C 5	38.4	4.0	2000	7 ADA71938	Ada71938 Rice gene
C 6	37.8	3.9	1413	4 ABL24039	Ab124039 Drosophil
C 7	35	3.6	1431	4 ABL07657	Ab107657 Drosophil
C 8	35	3.6	4200	4 ABL07656	Ab107656 Drosophil
C 9	34.8	3.6	1590	2 AAV19125	Aav19125 Nucleotid
C 10	34.8	3.6	1590	2 AAX25066	Aax25066 Brassica
C 11	34.8	3.6	1590	6 ABA97357	Aba97357 Brassica
C 12	33.8	3.5	7133	6 ABL58181	Ab158181 Murine St
C 13	33.6	3.5	1073	3 AAC39454	Aac39454 Arabidops
C 14	33.2	3.4	3780	4 ABL24038	Ab124038 Drosophil
C 15	33	3.4	348	6 ABL81387	Ab181387 Human ova
C 16	33	3.4	391	3 AAC09571	Aac09571 Human sec
C 17	33	3.4	481	4 AAL14686	Aal14686 Human bre
C 18	33	3.4	482	4 AAL23548	Aal23548 Human bre
C 19	33	3.4	493	4 AAS27046	Aas27046 cDNA enco
C 20	33	3.4	493	9 ADB3224	Adb3224 Human cDN
C 21	33	3.4	494	6 ABL79833	Ab179833 Human ova
C 22	33	3.4	532	3 AAC99042	Aac99042 Human pan
C 23	33	3.4	532	4 AAF72769	Aaf72769 Human pro

C 24	33	3.4	579	3 AAC77387	Aac77387 Human ORF
C 25	33	3.4	918	3 AAC76648	Aac76648 Human ORF
C 26	33	3.4	1022	6 ABK46152	Abk46152 cDNA enco
C 27	33	3.4	1097	3 AAA96893	Aaa96893 Nucleotid
C 28	33	3.4	1102	4 AAI58936	Aai58936 Human pol
C 29	33	3.4	1102	8 ADB48918	Adb48918 Novel hum
C 30	33	3.4	1129	3 AAC77841	Aac77841 Human can
C 31	33	3.4	1129	4 AAF72745	Aaf72745 Human pro
C 32	33	3.4	1129	4 AAF72770	Aaf72770 Human pro
C 33	33	3.4	1224	4 AAI60722	Aai60722 Human pol
C 34	32.8	3.4	7357	4 ABL18274	Ab118274 Drosophil
C 35	32.6	3.4	103599	4 ABX04971	Abx04971 S. cinnam
C 36	32.4	3.3	588	4 ABL25387	Ab125387 Drosophil
C 37	32.4	3.3	1167	6 RAD46542	Rad46542 Mouse dia
C 38	32.4	3.3	1881	6 ABK99846	Abk99846 Babesia c
C 39	32.4	3.3	2262	8 AAD56889	Aad56889 Mouse dia
C 40	32.4	3.3	4188	7 ABX13487	Abx13487 B. popill
C 41	32.4	3.3	8294	6 ABL34430	Ab134430 Human imm
C 42	32.2	3.3	345	5 ABV45947	Abv45947 Human pro
C 43	32.2	3.3	1707	4 ABL08339	Ab108339 Drosophil
C 44	32.2	3.3	6594	4 ABL08338	Ab108338 Drosophil
C 45	32.2	3.3	10982	4 AAS46264	Aas46264 DNA enco

## ALIGNMENTS

## RESULT 1

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX AC ADA71938;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 5263.

XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
XX gene; ds.

XX OS Oryza sativa.

XX PN WO2003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001; 2001WO-IB001105.

XX PR 22-JUN-2001; 2001WO-IB001105.

XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX DR WPI; 2003-1752590/17.

XX PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.

XX PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX CC The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance to a response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to



WP ABQ67197\_04 400001 510000  
 WP ABQ67197\_05 500001 610000  
 WP ABQ67197\_06 600001 710000  
 WP ABQ67197\_07 700001 810000  
 WP ABQ67197\_08 800001 910000  
 WP ABQ67197\_09 900001 1010000  
 WP ABQ67197\_10 1000001 1110000  
 WP ABQ67197\_11 1100001 1163020

Query Match 4.0%; Score 38.6; DB 6; Length 110000;  
 Best Local Similarity 61.4%; Pred. No. 1.5;  
 Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 453 GCGAGCGGAAGCTCCAGGCTGCGATGATCTCTGTCGATCTCCTCGTGGCGAGGAATT 512  
 DB 13671 GCAGCGCAATTGGCAGCGGTCGATGAATTTATCAAGTTGCGGTGCGCAGGAATC 13612

QY 513 TCTGCTCTCGCTCGCGGCGAGATCATTCGCTGAAGAAGA 553  
 DB 13611 GCTGCCCTAGCTTACGCTACAGCATCCATTCCTAAAGTAGA 13571

RESULT 5  
 ADA71938  
 ID ADA71938 standard; DNA; 2000 BP.  
 AC ADA71938;  
 DT 20-NOV-2003 (first entry)  
 XX Rice gene, SEQ ID 5263.  
 DE Plant; bacterial infection; fungal infection; viral infection; rice;  
 KW gene; ds.  
 XX Oryza sativa.  
 XX WO2003000898-A1.  
 XX 03-JAN-2003.  
 XX 22-JUN-2001; 2001WO-IB001105.  
 XX 22-JUN-2001; 2001WO-IB001105.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX WPI; 2003-175290/17.  
 XX Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX Claim 27; SEQ ID NO 5263; 899pp; English.  
 XX The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
 SQ

Query Match 4.0%; Score 38.4; DB 7; Length 2000;

Best Local Similarity 10.7%; Pred. No. 0.21;  
 Matches 42; Conservative 173; Mismatches 179; Indels 0; Gaps 0;

QY 6 CAGCAGTACGGCATCGACGAGCGGTAGCGAGCATCGATACGCGGGGCGCTGTCTAAA 65  
 DB 35 SRMRKMGSMKYKCKSSCGCKMTTTRKSKWYSASSASGRTGSKWSSGSYSKGKMKRY 94  
 QY 66 GCGTTTGGGTTTGACGCGCAACGATATAACGTCCTGAGCGCTGTGACATGCTCGCGCT 125  
 DB 95 KRKRWRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGRGR 154  
 QY 126 TGCGGTTTGAAGCGCGGAGATCGCTGGTGTGACTCGACCACTCCACGCTTGGCGTTCA 185  
 DB 155 SCRKKSGSGWKTCCRGARGSGSSGAKYKSGSKMWMSSCGRSGGRRSARSY 214  
 QY 186 GAGATCATGAGCTGTGTCGGCGAGCATGGGCGCAAGCCGATCGCCACCGAAGTCGAGA 245  
 DB 215 YGTSRKYGTGTYKMTYYSASRCMAYMTTSYSWACSSYTWCRSKRRSMWMMKMRKWSRS 274  
 QY 246 AGATTTCCGTAAGAAGAGCGCTTCGGCAGTGGGTGATAATGGCTTCGCCGCTACTTATGC 305  
 DB 275 YGWYSYKVMCTAYKYSYRWYMGWGGWGGATRYWGRGYMSRMMYKMYWYRG 334  
 QY 306 GGACAATTTAAACATCGCCACGAACTCGCAGCTCAACGAGTACTGTCTGTGACGAAGC 365  
 DB 335 YKGMKSGWAGRMMSRMSKRWKACYYMWRMWRMWRMWRMWRMWRMWRMWRMWRMWR 394  
 QY 366 ATTACAGCGCTAAGGCGCATATGGAACGGAGGC 399  
 DB 395 RYKRMRYGSRMSRCKRMRMCKRSGRAWMKGC 428

RESULT 6  
 ABL24039/c  
 ID ABL24039 standard; DNA; 1413 BP.  
 XX ABL24039;  
 XX 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23590.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 XX 11-JUL-2000; 2000US-00614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI; 2001-656860/75.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Claim 1; SEQ ID NO 23590; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention



CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1413 BP; 353 A; 434 C; 404 G; 222 T; 0 U; 0 Other;  
Query Match 3.9%; Score 37.8; DB 4; Length 1413;  
Best Local Similarity 45.0%; Pred. No. 0.28;  
Matches 141; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
QY 425 TTGATGAGCATCGAAGCGCTTAAAGCGGAGCGAGCTCTCAGGCTCGATGATCT 484  
DB 653 TTAGGAGTCCAGCTCTTGGCATAGCTAGCTCAATGGTCTCAGCCCGTATCCACC 594  
QY 485 CTGTGCTCTCCCTGGTGGGAGGAATTTCTGCTCTCCCTGCGGGGAGGAATCCATTGC 544  
DB 593 AGCCGGGAGGACAGGGTGTGATGATCTTCTGCTGGCGCTGGATGCTCAACTCCCGATGG 534  
QY 545 TGAAGAAGATGTCGAAGAGTTTGGCGCGAATTTGCACCTGCTCGACGGTGTCTTACC 604  
DB 533 TTCAATAGACTGGCTATCTGCTTCTGCTTCCGCTTCAGTCGGGATTCGAAGAGCAAGGC 474  
QY 605 TCGGGGATCTCGTGAAGCATTCCTATTTCGAACGAGAGATCTTTGAAGAGGTGCTCG 664  
DB 473 TCGGAGGTCAACGAGTTCTATCTGATCAGATTTCTCTTCTGTAACGATAGTGTCTGG 414  
QY 665 GCCTGACTTTGACCCCATGCTCGAATGCGGACCTTATTCTCCATCAGCAAGTTCGTG 724  
DB 413 ACCTGATGATACAGCTTTGGCCACTTCATGATGTGCTCGTCTGGAAGCGCACCTTG 354  
QY 725 GGTGGCTTTCTCT 737  
DB 353 GAGCGCAGCTCTCT 341  
RESULT 7  
ID ABL07657 standard; cDNA; 1431 BP.  
XX  
AC ABL07657;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 17453.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB63554.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 17453; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1431 BP; 390 A; 341 C; 342 G; 358 T; 0 U; 0 Other;  
Query Match 3.6%; Score 35; DB 4; Length 1431;  
Best Local Similarity 56.5%; Pred. No. 2.3;  
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 825 AATTTTGTAACTGCTTTGAAATCGTGAATCGATGCGTCAGTGGCTCCCTGACCTGGC 884  
DB 942 AATTTTGTGAACGACTTTCAAAATCCTCAGGCATTTGGTAGATATCTCAATAAGTTGGC 1001  
QY 885 TTGGAAACTCAGGCTGGGTAGCGGAGATCCAGACCATCTGATTCACCTTAAC 939  
DB 1002 GGATAACAAAAGCTGTACAATTCGTACCGCCACACAACTAAATCGCGGTAAAC 1056  
RESULT 8  
ID ABL07656/c  
XX ABL07656 standard; cDNA; 4200 BP.  
XX  
AC ABL07656;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 17450.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB63553.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 17450; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly

```
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 4200 BP; 1266 A; 838 C; 858 G; 1238 T; 0 U; 0 Other;

Query Match 3.6%; Score 35; DB 4; Length 4200;
Best Local Similarity 56.5%; Pred. No. 4.1;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 825 AATTTTGTAACTGCTTGAATAATCGTGAGTCGATCGTCAGTGGCTCCCTGAGCTGGC 884
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1490 AATTTTGTGACGACATTCGAAATCTCAGGATTTGGTAGAGTATCTCAATAAGTTGGC 1431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 885 TTGGAAACTGAGCGGTGGGTAGCGGAGATCCAGACCATCTGATTCACCTTAAC 939
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1430 GGATAACAAAAGCTGTACAATTCGTACCGCGCAACAACTAAATCGGCGTAC 1376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
AAV19125/c
ID AAV19125 standard; cDNA; 1590 BP.
XX AC
XX AAV19125;
XX DT 10-AUG-1998 (first entry)
XX DE Nucleotide sequence of beta cyclase cDNA clone.
XX KW Beta cyclase; epsilon cyclase; SSU/crtB fusion protein; carotenoid;
XX KW lycopene; lutein; zeaxanthin; canthaxanthin; vitamin; colourant; ss.
XX OS Brassica napus.
XX PN WO9806862-A1.
XX PD 19-FEB-1998.
XX PF 08-AUG-1997; 97WO-US014035.
XX PR 09-AUG-1996; 96US-0024145P.
XX PA (CALJ ) CALGENE INC.
XX PI Shewmaker CK;
XX WPI; 1998-159551/14.
XX PT Use of constructs comprising a carotenoid biosynthesis gene - for
XX PT producing plants and seeds having altered carotenoid levels, modified
XX PT fatty acid compositions or altered tocopherol levels.
XX PS Example 1B; Fig 11; 70pp; English.
XX CC This is the nucleotide sequence of the Brassica napus beta cyclase cDNA
XX CC clone. It was used in the method of the invention with the SSU/crtB
XX CC fusion protein (AAV19122), to create plants and seeds with altered
XX CC carotenoid levels. They can be used for the production of particular
XX CC carotenoids, e.g. alpha-carotene, beta-carotene, lycopene, lutein,
XX CC zeaxanthin, canthaxanthin, which can be used as supplements, particularly
XX CC vitamin supplements, as vegetable oil based food products and food
XX CC ingredients, as feed additives in animal feeds, as colourants or
XX CC therapeutic agents
XX SQ Sequence 1590 BP; 438 A; 440 C; 352 G; 360 T; 0 U; 0 Other;

Query Match 3.6%; Score 34.8; DB 2; Length 1590;
Best Local Similarity 52.0%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

QY 112 ACATTGCTCGGCTTGCCTTTGAAGCGGGAGATCGTGGTGTACTCGACCATCCA 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1180 ACGAGTTTGGGTGACGAGTTCGAGCCATGACTTGTAGACTGCTCGACCATCT 1121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 172 CG--CCTTGGCGTTCAGAGATCATGGACTGGTCCGCGAGCATTCGGGCCAAGCCGTACG 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

CC This is the nucleotide sequence of a Brassica napus lycopene beta cyclase
CC cDNA clone, which was obtained by PCR using primers designed from an
CC Arabidopsis beta cyclase gene. The xanthophyll content of seeds can be
CC altered by transforming cells of a host plant with at least one construct
CC containing: (i) a transcription initiation region of a gene
CC preferentially expressed in seeds; (ii) the sequence for a plastid
CC transit peptide; (iii) DNA from a carotenoid synthesis gene coding
CC region, and (iv) a transcription terminator. The transformed cells are
CC regenerated to plants and these, or their progeny, grown to produce
CC seeds. The method is especially used to increase the carotenoid content
CC in oilseed plants. The use of antisense DNA sequences prevent the
CC conversion of a selected precursor compound into the next carotenoid of
CC the carotenoid biosynthetic pathway. Thus, antisense lycopene beta-
CC cyclase genes can be used to increase production of lycopene
XX SQ Sequence 1590 BP; 438 A; 440 C; 352 G; 360 T; 0 U; 0 Other;

Query Match 3.6%; Score 34.8; DB 2; Length 1590;
Best Local Similarity 52.0%; Pred. No. 2.9;
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

QY 112 ACATTGCTCGGCTTGCCTTTGAAGCGGGAGATCGTGGTGTACTCGACCATCCA 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1180 ACGAGTTTGGGTGACGAGTTCGAGCCATGACTTGTAGACTGCTCGACCATCT 1121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 172 CG--CCTTGGCGTTCAGAGATCATGGACTGGTCCGCGAGCATTCGGGCCAAGCCGTACG 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1120 GGTCCGGCGCGTCTACATCAACGAGCGCTCCGAAAGGAGCTGTCGCGCTTACG 1061  
Qy 230 CCACCGGAAGTCGAGAGATTTCGTAAGAGAGCGCTTCGCGAGTGGTTCATATGCT 289  
Db 1060 GGAGAGTGAACCGTAACAGCTCAATCCAAAGATGCTTCGAGTGCATCCACACGGTG 1001  
Qy 290 TCGCGCGTACTTAATGC 305  
Db 1000 TTAGATTCATCAGGC 985

RESULT 11  
ID ABA97357 standard; cDNA; 1590 BP.  
XX AC ABA97357;  
XX  
XX 12-APR-2002 (first entry)  
XX Brassica napus beta cyclase cDNA clone.  
XX Carotenoid compound; oilseed; plant; vitamin A; ss; gene.  
XX Brassica napus.  
XX WO200188169-A2.  
XX 22-NOV-2001.  
XX 11-MAY-2001; 2001WO-US015264.  
XX 12-MAY-2000; 2000US-00570140.  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
XX Shewmaker CK;  
XX WPI; 2002-075318/10.  
XX  
XX Altering carotenoid level/content in oilseed plants, by transforming  
XX plants with construct having DNA sequence encoding a carotenoid  
XX biosynthesis gene, plastid transit peptide, transcription initiator and  
XX terminator.  
XX Disclosure; Page 72-73; 114pp; English.  
XX  
XX This invention relates to altering the carotenoid content or composition  
XX in seed from a host plant. This is achieved by transforming cells of host  
XX plant with a construct of operably linked components, a transcriptional  
XX initiation region from a gene expressed in a plant seed, a plastid  
XX transit peptide, a nucleic acid sequence encoding a carotenoid  
XX biosynthesis gene from eukaryotic source, and a transcriptional  
XX termination region. The method is useful for increasing carotenoid  
XX content in seed and endosperm of a host plant and altering carotenoid  
XX composition in a plant seed. The method is also useful for screening  
XX transformed corn seeds or transformed endosperms, where the transformed  
XX seed and endosperms are visually determined and selected based on yellow,  
XX orange or red colour as result of the increased carotenoid content. The  
XX transformed seeds provide a source of modified oils and the oil extracted  
XX from the seeds is useful as a food colourant, or as a food oil with high  
XX alpha and beta-carotene levels for prevention of vitamin A deficiency  
XX which can result in night blindness. The level of lutein is also  
XX increased in seeds of transformed plants. The modified oil obtained from  
XX the transformed seeds are more oxidatively stable than naturally  
XX occurring oils and are low saturate, high oleic and low linolenic. This  
XX sequence represents the nucleotide sequence of the Brassica napus beta  
XX cyclase cDNA clone  
XX  
SQ Sequence 1590 BP; 438 A; 440 C; 352 G; 360 T; 0 U; 0 Other;  
Query Match 3.6%; Score 34.8; DB 6; Length 1590;  
Best Local Similarity 52.0%; Pred. No. 2.9;  
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

Qy 112 ACATTGCTCGCGCTTGCAGTTTGAAGCCGGGAGATCGCTGGTTGACTCGACCACTCCA 171  
Db 1180 ACGAGTTTGGTTGACGAGTTTCGAGCCATGACTTGTAGACTGCTCGACACCACT 1121  
Qy 172 CG--CCTTGGCGTTTCAGAGATCATGACTGTCGGGCGAGCANTGGGCCAAGCGGTACG 229  
Db 1120 GGTCCGGCGCGTCTACATCAACGAGCGCTCCGAAAGGAGCTGTCGCGCGCTTACG 1061  
Qy 230 CCACCGGAAGTCGAGAGATTTCGTAAGAGAGCGCTTCGCGAGTGGTTCATATGCT 289  
Db 1060 GGAGAGTGAACCGTAACAGCTCAATCCAAAGATGCTTCAGAGTGCATCCACACGGTG 1001  
Qy 290 TCGCGCGTACTTAATGC 305  
Db 1000 TTAGATTCATCAGGC 985

RESULT 12  
ID ABL58181/c  
XX ABL58181 standard; DNA; 7133 BP.  
XX AC ABL58181;  
XX  
XX 05-AUG-2002 (first entry)  
XX Murine Stable Tubulin Only Polypeptide, STOP, gene fragment #1.  
XX STOP; neuroleptic; tranquiliser; antidepressant; transgenic animal;  
XX Stable Tubulin Only Polypeptide; drug screening; schizophrenia; gene;  
XX schizoaffective disorder; anxiety; paranoia; depression; murine; ds.  
XX Mus musculus.  
XX WO200241691-A2.  
XX 30-MAY-2002.  
XX 23-NOV-2001; 2001WO-FR003701.  
XX 24-NOV-2000; 2000FR-00015240.  
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
XX (INRM ) INST NAT SANTE & RECH MEDICALE.  
XX Andrieux A, Job D, Denarier E, Bosc C, Vernet M;  
XX WPI; 2002-463598/49.  
XX Recombinant mammal useful in screening for agents to treat e.g.  
XX schizophrenia, comprises a modified allele for stable tubulin only  
XX polypeptide (STOP) protein.  
XX Example 1; Fig 2; 57pp; French.  
XX The present invention relates to a non-human recombinant animal (A),  
XX which contains a modified allele of the gene for STOP (Stable Tubulin  
XX Only Polypeptide) protein. (A), or their cell extracts or organ slices  
XX (preferably neuronal or cerebral) are used to screen for drugs useful in  
XX treatment of schizophrenia or schizoaffective disorders that have an  
XX anxiety, paranoia or depressive component. The present sequence is a gene  
XX fragment of the murine STOP gene, used to illustrate the invention  
XX  
SQ Sequence 7133 BP; 1765 A; 1941 C; 1718 G; 1709 T; 0 U; 0 Other;  
Query Match 3.5%; Score 33.8; DB 6; Length 7133;  
Best Local Similarity 51.7%; Pred. No. 14;  
Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 132 TTTGAAGCCGGGAGATCGCTGGTTCGACTCGACCACTCCACGCTTGGCGTTTCAGAGAT 191  
Db 4535 TTTCTGTCGCCGAGCACCACCGCGCTCTCTCCAGCACCTCCCGGGTCCCGGGCATCAGCAGAA 4476

```
QY      192 CATGGACTGTCCGGCGAGCATTCGGCCAAAGCCCTACGCCACCGGAAGTCGAGAAGATTT 251
Db      4475 AGTTCATGGCGCGCGCTCTTGCTCTGAGCCGACGCTTGGCGACCCCGAGAACAGGT 4416
QY      252 CCGTAAGAACGCGTTCGGCGAGTGGTTG 280
Db      4415 TGGGAAGCTCGGAAGTCCGAGGATTTG 4387

RESULT 13
ID      AAC39454
XX      AAC39454 standard; DNA; 1073 BP.
AC      AAC39454;
DT      17-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana DNA fragment SEQ ID NO: 24688.
XX
KW      Hybridisation assay; genetic mapping; gene expression control;
KW      protein identification; signal transduction pathway; metabolic pathway;
KW      promoter; termination sequence; ss.
XX
OS      Arabidopsis thaliana.
XX
PN      EP1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-00301439.
XX
PR      25-FEB-1999; 99US-0121825P.
PR      05-MAR-1999; 99US-0123180P.
PR      09-MAR-1999; 99US-0123548P.
PR      23-MAR-1999; 99US-0125788P.
PR      25-MAR-1999; 99US-0126264P.
PR      29-MAR-1999; 99US-0126785P.
PR      01-APR-1999; 99US-0127462P.
PR      06-APR-1999; 99US-0128234P.
PR      08-APR-1999; 99US-0128714P.
PR      16-APR-1999; 99US-0129845P.
PR      19-APR-1999; 99US-0130077P.
PR      21-APR-1999; 99US-0130449P.
PR      23-APR-1999; 99US-0130510P.
PR      23-APR-1999; 99US-0130891P.
PR      28-APR-1999; 99US-0131449P.
PR      30-APR-1999; 99US-0132048P.
PR      30-APR-1999; 99US-0132407P.
PR      04-MAY-1999; 99US-0132484P.
PR      05-MAY-1999; 99US-0132485P.
PR      06-MAY-1999; 99US-0132486P.
PR      06-MAY-1999; 99US-0132487P.
PR      07-MAY-1999; 99US-0132863P.
PR      11-MAY-1999; 99US-0134256P.
PR      14-MAY-1999; 99US-0134218P.
PR      14-MAY-1999; 99US-0134219P.
PR      14-MAY-1999; 99US-0134370P.
PR      18-MAY-1999; 99US-0134768P.
PR      19-MAY-1999; 99US-0134941P.
PR      20-MAY-1999; 99US-0135124P.
PR      21-MAY-1999; 99US-0135353P.
PR      24-MAY-1999; 99US-0135629P.
PR      25-MAY-1999; 99US-0136021P.
PR      27-MAY-1999; 99US-0136392P.
PR      28-MAY-1999; 99US-0136782P.
PR      01-JUN-1999; 99US-0137222P.
PR      03-JUN-1999; 99US-0137528P.
PR      04-JUN-1999; 99US-0137502P.
PR      07-JUN-1999; 99US-0137724P.
PR      08-JUN-1999; 99US-0138094P.
PR      10-JUN-1999; 99US-0138540P.
PR      10-JUN-1999; 99US-0138847P.
PR      14-JUN-1999; 99US-0139119P.
PR      16-JUN-1999; 99US-0139452P.
PR      16-JUN-1999; 99US-0139453P.
PR      17-JUN-1999; 99US-0139492P.
PR      18-JUN-1999; 99US-0139454P.
PR      18-JUN-1999; 99US-0139455P.
PR      18-JUN-1999; 99US-0139456P.
PR      18-JUN-1999; 99US-0139457P.
PR      18-JUN-1999; 99US-0139458P.
PR      18-JUN-1999; 99US-0139459P.
PR      18-JUN-1999; 99US-0139460P.
PR      18-JUN-1999; 99US-0139461P.
PR      18-JUN-1999; 99US-0139462P.
PR      18-JUN-1999; 99US-0139463P.
PR      18-JUN-1999; 99US-0139750P.
PR      18-JUN-1999; 99US-0139763P.
PR      21-JUN-1999; 99US-0139817P.
PR      22-JUN-1999; 99US-0139899P.
PR      23-JUN-1999; 99US-0140353P.
PR      23-JUN-1999; 99US-0140354P.
PR      24-JUN-1999; 99US-0140695P.
PR      28-JUN-1999; 99US-0140823P.
PR      29-JUN-1999; 99US-0140991P.
PR      30-JUL-1999; 99US-0141287P.
PR      01-JUL-1999; 99US-0141842P.
PR      02-JUL-1999; 99US-0142154P.
PR      06-JUL-1999; 99US-0142055P.
PR      06-JUL-1999; 99US-0142390P.
PR      08-JUL-1999; 99US-0142803P.
PR      09-JUL-1999; 99US-0142920P.
PR      12-JUL-1999; 99US-0142977P.
PR      13-JUL-1999; 99US-0143542P.
PR      14-JUL-1999; 99US-0143624P.
PR      15-JUL-1999; 99US-0144005P.
PR      16-JUL-1999; 99US-0144085P.
PR      19-JUL-1999; 99US-0144325P.
PR      19-JUL-1999; 99US-0144331P.
PR      19-JUL-1999; 99US-0144332P.
PR      19-JUL-1999; 99US-0144333P.
PR      19-JUL-1999; 99US-0144334P.
PR      20-JUL-1999; 99US-0144352P.
PR      20-JUL-1999; 99US-0144632P.
PR      20-JUL-1999; 99US-0144884P.
PR      21-JUL-1999; 99US-0144814P.
PR      21-JUL-1999; 99US-0145086P.
PR      21-JUL-1999; 99US-0145088P.
PR      22-JUL-1999; 99US-0145085P.
PR      22-JUL-1999; 99US-0145087P.
PR      22-JUL-1999; 99US-0145089P.
PR      22-JUL-1999; 99US-0145192P.
PR      23-JUL-1999; 99US-0145145P.
PR      23-JUL-1999; 99US-0145218P.
PR      23-JUL-1999; 99US-0145224P.
PR      26-JUL-1999; 99US-0145276P.
PR      27-JUL-1999; 99US-0145913P.
PR      27-JUL-1999; 99US-0145918P.
PR      27-JUL-1999; 99US-0145919P.
PR      28-JUL-1999; 99US-0145951P.
PR      02-AUG-1999; 99US-0146386P.
PR      02-AUG-1999; 99US-0146388P.
PR      03-AUG-1999; 99US-0146389P.
PR      04-AUG-1999; 99US-0147038P.
PR      04-AUG-1999; 99US-0147204P.
PR      05-AUG-1999; 99US-0147302P.
PR      05-AUG-1999; 99US-0147192P.
PR      06-AUG-1999; 99US-0147260P.
PR      06-AUG-1999; 99US-0147303P.
PR      09-AUG-1999; 99US-0147416P.
PR      09-AUG-1999; 99US-0147493P.
PR      10-AUG-1999; 99US-0147935P.
PR      10-AUG-1999; 99US-0148171P.
```

PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157863P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161933P.  
PR 29-OCT-1999; 99US-0162142P.  
  
Query Match 3.5%; Score 33.6; DB 3; Length 1073;  
Best Local Similarity 50.0%; Pred. No. 5.8;  
Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 396 AGCTTCGAGGAATCTCTTGTAGTCTTCTTGTAGTGAAGCATCGAAGCGGTTAAAGCGG 455  
Db 320 AGTCTTCAATGCTGCTCTTCTTGTAGTCTTCTTGTAGTGAAGCATCGAAGCGGTTAAAGCGG 379

QY 456 AGCGAAGCTCTCCAGGCTCGGATGATCTCTGTGATCTCCTCGTGGCGGAGGAATTTCT 515  
Db 380 TAGGCAATTTCTTTCTTCCTCGCTTAGTATTGAAGATTATAACGATCTGTGTAGAGCTGT 439  
QY 516 GCTCTCGCTGCCGGGAGAGATCCATTGCTGAAGAGATGCTCAAGA 563  
Db 440 GGTGCGTGTGAAGCTCAGTATGCTTTGCTGAAGAACTTTGTTGATGA 487

RESULT 14  
ABL24038/c  
ID ABL24038 standard; DNA; 3780 BP.  
XX ABL24038;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23587.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEXE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

PS Claim 1; SEQ ID NO 23587; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3780 BP; 1059 A; 951 C; 871 G; 899 T; 0 U; 0 Other;

Query Match 3.4%; Score 33.2; DB 4; Length 3780;  
Best Local Similarity 45.9%; Pred. No. 15;  
Matches 113; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
QY 425 TTGATGAAGCATCGAAGCGGTTAAAGCGGAGCGGAGCTCTCCAGGCTCGATGATCT 484  
Db 1741 TTCAGGAGTCCAGCTCGTTGGCATAGTAGTTCATGCTCCAGCCGATCCACC 1682

QY 485 CTGTGATCTCCCTGGTGGCGAGGAATTTCTGCTCGCTGCCGGGAGATCCATTGC 544  
Db 1681 AGCCGGAGGACAGGGGTGTCGATGATCTTCTGTCGGCTGGATGCTCAATCCCGATGG 1622  
QY 545 TGAAGAGATGTCGAAGAGATTGTGCCGCGATTTCACCTCGCTCGAGGTCTCTTACC 604

Db 1621 TTCATAGACTGGCTATCTGCTTCTGCTTCCGCTTCACTCGGATTCACAGAGCAAGAGC 1562  
QY 605 TCGGGATACTCGTGGAAAGCATTCCTATTTCGAACGAGAGATCTTTGAAGAGGTGCTCG 664  
Db 1561 TCGGAGGTCAACGAGTTCATCTGATCAGATTCTCTTCTGTAACGATAGTGTGCTGG 1502  
QY 665 GCCTGA 670  
Db 1501 ACCTGA 1496

## RESULT 15

ABL81387/c

ID ABL81387 standard; cDNA; 348 BP.

XX

AC ABL81387;

XX

XX 17-MAY-2002 (first entry)

XX

XX Human ovarian cancer related cDNA clone SEQ ID NO:4365.

DE

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX

XX Homo sapiens.

OS

XX WO200192581-A2.

PN

XX 06-DEC-2001.

XX

XX 29-MAY-2001; 2001WO-US017756.

XX

XX 26-MAY-2000; 2000US-0207484P.

PR

XX (CORI-) CORIXA CORP.

XX

XX Algate PA, Harlocker SL, Jones R;

PI

XX WPI; 2002-122075/16.

XX

XX Composition for therapy and diagnosis of ovarian cancer comprising

PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding

PT polypeptide, antibody specific to polypeptide or T cell expressing

PT polypeptide.

XX

XX Claim 1; SEQ ID NO 4365; 489pp; English.

PS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 105 CGCTGTGACATTGCTCGCGCTTCCGCTTGAAGCCGGAGATCGCTGGTTGACTCGAC 164  
Db 196 CGCATGGAGGTGCAAGGTGCTTCGGGACAGAGCGACAAATCTGTGTGTGGCTCAAT 137  
QY 165 CACTCCACGCCCTTGGCGTTTCAAGAAGATCATGGACTGGTCCGGCGAGCATTTGGG 217  
Db 136 GCCCTCAACCCCTCAGCTCTCAAGAAGATCTGTCTGCCCTCCTCTCTCATTTGGG 84

Search completed: October 2, 2004, 00:28:19

Job time : 419.286 secs

The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (II) of a ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL77023 to ABL87934. (III) encoding (II) having a sequence (S2), a T cell population of (II), or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (IV) is detected preferably by polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising contacting T cells with (III) or (II). (III) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a suitable library e.g., a tumour cDNA library using well known techniques

Sequence 348 BP; 82 A; 102 C; 94 G; 70 T; 0 U; 0 Other;

Query Match

3.4%; Score 33; DB 6; Length 348;

Best Local Similarity 55.8%; Pred. No. 5;

Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 23:40:18 ; Search time 77.5043 Seconds  
(without alignments)  
6959.771 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972

Sequence: 1 gtgaacagcagtgacggcat.....ttggccgctacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	34.8	3.6	1590	US-08-908-758-4	Sequence 4, Appli
C 2	34.8	3.6	1590	US-09-023-587A-4	Sequence 4, Appli
C 3	33.8	3.5	1227	US-09-489-039A-2533	Sequence 2533, Ap
C 4	33	3.4	1102	US-09-620-312D-828	Sequence 828, Ap
C 5	32.8	3.4	1020	US-09-252-991A-9936	Sequence 9936, Ap
C 6	32.8	3.4	1728	US-09-252-991A-9757	Sequence 9757, Ap
C 7	32.8	3.4	1824	US-09-252-991A-10076	Sequence 10076, A
C 8	32	3.3	505	US-09-621-976-15639	Sequence 15639, A
C 9	32	3.3	2898	US-09-489-039A-1679	Sequence 1679, Ap
C 10	32	3.3	5633	US-09-023-655-1490	Sequence 1490, Ap
C 11	31.8	3.3	36941	US-08-311-731A-130	Sequence 130, Ap
C 12	31.6	3.3	3291	US-08-506-296B-3	Sequence 3, Appli
C 13	31.2	3.2	1680	US-09-375-975-39	Sequence 39, Appli
C 14	31.2	3.2	15872	US-09-105-537-1	Sequence 1, Appli
C 15	31.2	3.2	15872	US-09-091-609-1	Sequence 1, Appli
C 16	31.2	3.2	15872	US-09-091-609-3	Sequence 3, Appli
C 17	31.2	3.2	77536	US-09-410-551B-1	Sequence 2, Appli
C 18	31	3.2	2187	US-09-127-219B-2	Sequence 2, Appli
C 19	30.8	3.2	1584	US-09-328-352-775	Sequence 775, Ap
C 20	30.6	3.1	1095	US-09-891-641-46	Sequence 46, Appl
C 21	30.6	3.1	2496	US-09-252-991A-8330	Sequence 8330, Ap
C 22	30.6	3.1	89047	US-09-596-002-34	Sequence 34, Appli
C 23	30.6	3.1	4403765	US-09-103-840A-2	Sequence 2, Appli
C 24	30.6	3.1	4411529	US-09-103-840A-1	Sequence 1, Appli
C 25	30.4	3.1	1802	US-09-221-017B-1019	Sequence 1019, Ap
C 26	30.4	3.1	9686	US-09-221-017B-1003	Sequence 1003, Ap
C 27	30.2	3.1	406	US-08-318-193-52	Sequence 52, Appli

C 28	30.2	3.1	424	2	US-08-765-783A-28	Sequence 28, Appl
C 29	30.2	3.1	424	3	US-08-921-100-28	Sequence 28, Appl
C 30	30.2	3.1	424	3	US-08-880-142-28	Sequence 28, Appl
C 31	30.2	3.1	424	3	US-08-902-201-28	Sequence 28, Appl
C 32	30.2	3.1	424	3	US-09-416-557-28	Sequence 28, Appl
C 33	30.2	3.1	1782	4	US-09-252-991A-15714	Sequence 15714, A
C 34	30.2	3.1	2208	4	US-09-252-991A-15601	Sequence 15601, A
C 35	30.2	3.1	3132	4	US-09-252-991A-11872	Sequence 11872, A
C 36	30	3.1	1002	4	US-09-724-623-28	Sequence 28, Appl
C 37	30	3.1	1023	4	US-09-252-991A-11508	Sequence 11508, A
C 38	30	3.1	1464	4	US-09-252-991A-11733	Sequence 11733, A
C 39	30	3.1	1629	4	US-09-620-312D-715	Sequence 715, App
C 40	30	3.1	1981	4	US-09-132-118-1	Sequence 1, Appli
C 41	30	3.1	2016	3	US-08-444-005-16	Sequence 16, Appli
C 42	30	3.1	2137	1	US-09-161-443-1	Sequence 1, Appli
C 43	30	3.1	2617	3	US-09-023-655-914	Sequence 914, App
C 44	30	3.1	2617	4	US-09-133-944-2	Sequence 2, Appli
C 45	30	3.1	9687	3		

#### ALIGNMENTS

#### RESULT 1

US-08-908-758-4/c

; Sequence 4, Application US/08908758

; Patent No. 6429356

; GENERAL INFORMATION:

; APPLICANT: Calgene, Inc.

; TITLE OF INVENTION: METHODS FOR PRODUCING CAROTENOID COMPOUNDS AND SPECIALTY OIL

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Calgene, Inc.

; STREET: 1920 Fifth Street

; CITY: Davis

; STATE: CA

; COUNTRY: USA

; ZIP: 95616

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB

; COMPUTER: IBM PC

; OPERATING SYSTEM: Windows NT 4.0

; SOFTWARE: Microsoft Word For Windows 7.0a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/908,758

; FILING DATE: 8-8-97

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/024,145

; FILING DATE: 8-9-96

; ATTORNEY/AGENT INFORMATION:

; NAME: Donna E. Scherer

; REGISTRATION NUMBER: 34,719

; REFERENCE/DOCKET NUMBER: 36,924

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (916) 753-6313

; TELEFAX: (916) 753-1510

; INFORMATION FOR SEQ ID NO: 4 :

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1590 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

US-08-908-758-4

Query Match

Best Local Similarity 3.6%; Score 34.8; DB 4; Length 1590;

Matches 102; Conservative 52.0%; Pred. No. 0.41;

Matches 102; Conservative 92; Indels 2; Gaps 1;

QY

112 ACATGCTCGCGCTTGGCTTGAAGCGGAGATCGCTGGTGAATCGACCTGCACTCA 171



Db 1180 ACGAGTTGGTTGACGAGTTGAGACCATGAGCTGCTAGACTGCTCGACACCACT 1121  
Qy 172 CG--CCTTGGCGTTTCAAGAATCATGTGCTCGGCGAGCATTTGGGCCAAGCGGTACG 229  
Db 1120 GGTCCGGCGCGTCTGTCTACATCAACGAGCGCTCCGAAAGAGACCTGTCCGGCGCTTACG 1061  
Qy 230 CCACGGGAAGTCGAGAAGATTTCGTAAAGAAGACGCTTCGCGAGTGGGTTGATATGGCT 289  
Db 1060 GGAGAGTGAACCGTAAACAGCTCAATCAAGATGCTTCAGAAGTGCATCAACCAACGCTG 1001  
Qy 290 TCGCGGTACTTAATGC 305  
Db 1000 TTAGATTCATCAGGC 985

RESULT 2  
US-09-023-587A-4/c  
; Sequence 4, Application US/09023587A  
; Patent No. 6653530  
; GENERAL INFORMATION:  
; APPLICANT: Shewmaker, Christine K.  
; APPLICANT: Bhat, Ganesh B.  
; APPLICANT: Venkatramesh, Mylavaram  
; APPLICANT: Rangwala, Shaikat H.  
; APPLICANT: Kishore, Ganesh M.  
; APPLICANT: Boddupalli, Sekhar S.  
; TITLE OF INVENTION: Methods for Producing Carotenoid Compounds, Tocopherol Compounds,  
; TITLE OF INVENTION: Specialty Oils in Plant Seeds  
; FILE REFERENCE: 16516.122  
; CURRENT APPLICATION NUMBER: US/09/023,587A  
; CURRENT FILING DATE: 1998-02-13  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1590  
; TYPE: DNA  
; ORGANISM: Brassica napus  
US-09-023-587A-4

Query Match 3.6%; Score 34.8; DB 4; Length 1590;  
Best Local Similarity 52.0%; Pred. No. 0.41;  
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;  
Qy 112 ACATTGCTCGCGTTCCCGTTTGAAGCGCGGAGATCGCTGGTTGACTCGACCACTCA 171  
Db 1180 ACGAGTTTGGGTTGACGAGTTTCAAGCCATGAGCTTGTAGACTGCTCGACCACT 1121  
Qy 172 CG--CCTTGGCGTTTCAAGAATCATGTGCTCGGCGAGCATTTGGGCCAAGCGGTACG 229  
Db 1120 GGTCCGGCGCGTCTGTCTACATCAACGAGCGCTCCGAAAGAGACCTGTCCGGCGCTTACG 1061  
Qy 230 CCACCGGAAGTCGAGAAGATTTCGTAAAGAAGACGCTTCGCGAGTGGGTTGATATGGCT 289  
Db 1060 GGAGAGTGAACCGTAAACAGCTCAATCAAGATGCTTCAGAAGTGCATCAACCAACGCTG 1001  
Qy 290 TCGCGGTACTTAATGC 305  
Db 1000 TTAGATTCATCAGGC 985

RESULT 3  
US-09-489-039A-2533  
; Sequence 2533, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Bleton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2533  
; LENGTH: 1227  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2533  
Query Match 3.5%; Score 33.8; DB 4; Length 1227;  
Best Local Similarity 50.2%; Pred. No. 0.75;  
Matches 135; Conservative 0; Mismatches 132; Indels 2; Gaps 2;  
Qy 1 GTGAACAGCAGTCAGCGCATCGACGGAACGGTAGGAGCATCGATACTGCGGGGGGCTG 60  
Db 961 GTGAGTCCATTGATTTAACTCAACGAGGCTTCGCGCTCAGCGCTGCGGCTG 1020  
Qy 61 CTAAAGCGTTTGGGTTTACGCGCAACGATATACGTCCTGAGCTGTGACATTGCTC 120  
Db 1021 ATGAACGATCTGCGTCTCGACCCGCGCGTTTAACTCAACGAGGCTTCGCGCTCAGCGCTGCGGCTG 1080  
Qy 121 GCGCTTGGCGGTTTGAAGCCGCGGAGATCGCTGGGTTGACTCGACCACTCCAGCGCTTGGC 180  
Db 1081 G-GCCACCGGTTAGCGCGCAGCGGTGCGATTCTGCGGTCAACTGATGATGAATGCG 1139  
Qy 181 GTTCAGAAAGATCATGACTGCTGCGCGAGCATTTGGGCCAAGCCGTACGCCACCGGAAGT 240  
Db 1140 GCGCAGCGCGCAGCAGCTGCTGCTGTCACCA-TGTGCATCGCGCGCGCGGCGGATTT 1198  
Qy 241 CGAGAAGATTTCGTTAAGAAGCGTTTCG 269  
Db 1199 CTATGCTGTTGAGCGCGAAGCGGCTTAG 1227

RESULT 4  
US-09-620-312D-828/c  
; Sequence 828, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aiding J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Danrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 828  
; LENGTH: 1102  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (234)..(875)

US-09-620-312D-828

	Query Match.	3.4%;	Score 33;	DB 4;	Length 1102;
	Best Local Similarity	55.8%;	Pred. No. 1.3;		
	Matches	63;	Conservative	0;	Mismatches 50; Indels 0; Gaps 0;
QY	105	CGCTGTGCATTTGCTCCGGCTTCCGGTTTGAACCGGGAGATCGCTGGTTGACTCGAC	164		
Db	225	CGCATGGAGGTTGCAGGTGTCTTCGGGACAGAGCGACAATCTGTGTGTGGCTCAAT	166		
QY	165	CATCCACGCCITGGCGCTTCAGAGATCATGGACTGTCCGGCGAGCATTTGGG	217		
Db	165	GCCTCAACCCCTCAGCTCTCAAAAGATCTGTCTCGCCCTCTCTCATTTGGG	113		

## RESULT 5

```

US-09-252-991A-9936
; Sequence 9936, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

US-09-252-991A-9936

Query Match 3.4%; Score 32.8; DB 4; Length 1020;  
Best Local Similarity 45.4%; Pred. No. 1.4;  
Matches 118; Conservative 0; Mismatches 142; Indels 0

95	AGCTCCGTAGCGCTGTGACATGTCTCGCGCTTGCCGCTTTGAAGCCGGAGATCGCTGGG	154
318	AGGTCGCGTATACGCGGAACCCGACTACTACGCGCGCAAGCCGCGATCGACAACCTGG	377
155	TTGACTCGAACACTCCACGCTTGGCGCTTCAGAAAGATCATGGACTGGTCGGCGGAGCATTT	214
378	TGTTTCGCATTACCCCTCGATCCCACTGGCGCATGCAGAAGTCTCGTCCGCGGAGTGGCC	437
215	GGGCCAAGCGCTACGCCACCGGAAGTCGAGAAGATTTCGCTAAGAAAGACGCTTCGCGCAGT	274
438	AGGTCCTCGCTGTATCCCGAAACCCGAGGACCTGCGCGCGCTGAAGCAGGACCCGAACTGG	497
275	GGGTTTGATATGCTTCGCGCTACTTAAATCGGACAAATTAAACATCGCGCAGAACTCGC	334
498	CGGTGGACGAGATCGATGCCCTGCTGACCACTTACATTGCCATTCACACCCAGCAACAAGC	557
335	AGCTCAACGAGTACTGCTTG	354
558	CGCTCGACGACCCGCGCGTG	577

## RESULT 6

US-09-252-991A-9757  
; Sequence 9757, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AGRUNOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9757
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9757

```

Query Match	3.4%;	Score 32.8;	DB 4;	Length 1728;
Best Local Similarity	45.4%;	Pred. No. 2;		
Matches 118; Conservative	0;	Mismatches 142;	Indels 0	

QY	95	ACGTCGTAGCGCTGTGACATTGCTCGCGTTTGGCCGTTTGAAGCCGGGAGATCGCTGGG	154
Db	785	AGGTCTCGCTATACGGCGAAACCGGACTACTACGCGCGCAAGCCGCGATCGACAACTGG	844
QY	155	TTGACTCGACACTCCACGCGCTTGGCGTTTCAGAAAGATCATGACTGGTCCGGCGGAGCAATT	214
Db	845	TGTTGGCCATTACCTCGATCCCACTGGTCGCTGCGATGCGAGAGGTCCGTGCCGCGGAGTGCC	904
QY	215	GGGCCAAGCGGTACGCAACCGGAAGTCGAGAAGATTTCCGTAAAGAAGACGGTTTCGGCAGT	274
Db	905	AGGTCCTCGTGTATACCCGAAACCGGAGACGTGCGCGCCTGAACAGGACCCGAACCTGG	964
QY	275	GGGTTGATATGCTTCGCCGTACTTAATCGGAGCAATTTAAACATCGCCACGAACTCGC	334
Db	965	CGGTGGACGAGATCGATGCGCCCTGTGTACCACCTACATTGCGATCAACACCCAGCAACAGC	1024
QY	335	AGCTCAACGAGTACTGCTTG	354
Db	1025	CGCTCGACGACCCGCGCGTG	1044

## RESULT, T 7

```

US-09-252-991A-10076/c
; Sequence 10076, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10076
; LENGTH: 1824
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10076

```

Query Match	3.4%;	Score 32.8;	DB 4;	Length 1824;
Best Local Similarity	45.4%;	Pred. No. 2;		
Matches 118;	Conservative	0;	Mismatches 142;	Indels 0;
				Gaps 0;

95	ACGTCGTAAGCGCTGTGACATTTGCTCGCGCTTGCCTGTTTGAAGCCGGAGATCGCTGGG	154
1214	AGGTCGGCTATACGGCGAACCCGGACTACTACGCGGCAAGCCCGCATCGACAACCTGG	1155
155	TTGAATCGACCACTCAAGCGCTTGGCTTCAAGAATCATGGAATGCTGGCGGAGCAATT	214
1154	TGTTGCCATTACCTTCGATCCCAAGTGGCATGCAGAAGGTCGTCGCGCGAGTGCC	1095
215	GGGCCAAGCGGTACGCCACCGGAAGTCGAGAAGATTTCCGTAAAGAGACGTTCCGCAGT	274
1094	AGGCTCTCGCTGTACCCGAAACCGGAGGAGCTGCGCGCTCAAGCAGGACCCGAACTGG	1035

Qy 275 GGGTTGATAGGCTTCGCGTACTTAAATCGGACAAATTTAAACATCGCCACGAACTCGC 334  
Db 1034 CGGTGGACGAGATCGATGCCCTGCTGACCACTACATTGCCATCAACACCCAGCAGC 975

Qy 335 AGCTCAACGAGTACTGCTTG 354  
Db 974 CGCTCGACGACCGCGCTG 955

RESULT 8  
US-09-621-976-15639  
; Sequence 15639, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 15639  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-15639

Query Match 3.3%; Score 32; DB 4; Length 505;  
Best Local Similarity 10.8%; Pred. No. 1.8;  
Matches 23; Conservative 111; Mismatches 96; Indels 0; Gaps 0;  
Qy 39 CATCGATACGCGCGGCGCTGCTAAAGCGTTTGGTTGACGCGCAAGATATAACGT 98  
Db 2 MAYMSRYRTSSKRYGKYSYSSRWMSMYAWGRKYGTSGRCGSGRWMSKMGY 61  
Qy 99 CGGTAGCGTGTGACATCTCGCGTTCGCGTTTGAAGCGGAGATCGCTGGTTGA 158  
Db 62 RYSYGYKWSKWKYKSGKMGTSKSTRKYRTYTSKCRKTKYRGSWKNRWRKR 121  
Qy 159 CTCGACCACTCCAGCGCTTGGCGTTCAGAAATCATGAGTCTCGCGGAGCATTTGGC 218  
Db 122 KMYYMKYCASCYSYRCKRYGTMTGVRGWGCKRCKSKSTRYMYTRYWMTGAC 181  
Qy 219 CAAGCGGTAGCCACCGGAGTGCAGAGATTTCCGTAAAGACGCTTC 268  
Db 182 YGSKMGCKSGRSKYGYGKWKYGYKTYMCTSKYKSKMSYKSKSMCYTM 231

RESULT 9  
US-09-489-039A-1679  
; Sequence 1679, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 1679  
; LENGTH: 2898  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-1679

Query Match 3.3%; Score 32; DB 4; Length 2898;  
Best Local Similarity 48.4%; Pred. No. 5;

Matches 89; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
Qy 707 TCCATGACGAAGTTTCGTGGTGGCTTTTCCCTTATGAGGCGCGTGAAGTAAAGTCCGT 766  
Db 1337 TCTATGAGCTTCAGTGTGGGTTCTGGCGCAGTATGAAGCGTGTGAATAAAAAACGG 1396  
Qy 767 TTGATGAGGAGCGGATCGAGCCTGCAAGAGCTATTGCTTACACCTTCAGGGGTCTAA 826  
Db 1397 TCTGGGAAGCGCGCAAGAGGAGCATGGGGGAAGAGATTGAAACCCCTCTTCGATCGGCTGA 1456  
Qy 827 TTTTGTAACTGCTTTGAAATCGTGAGTCGATCGCTCAGTGGCTCCCTCAGCTGGCTT 886  
Db 1457 TCTTTGATATCATGACTCCAGTAGTCTGCGAGATGCTGCTGATCGCTCATCAGGTT 1516  
Qy 887 GGGG 890  
Db 1517 GGGG 1520

RESULT 10  
US-09-023-655-1490  
; Sequence 1490, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1490:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5633 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: 9862374  
; US-09-023-655-1490

Query Match 3.3%; Score 32; DB 4; Length 5633;  
Best Local Similarity 52.2%; Pred. No. 7.3;  
Matches 71; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
Qy 823 CTAATTTTGTAAAGCTTTGAAAAATCGTGAGTCGATGCGTCAAGTGGCTCCCTCAGCTG 882

Db 3 CGAGTGTGGTTCACCTGTGGAGACAGCGGTGGCGGAGTGGGTCTCCAGGGCTCTGGGCTG 62  
QY 883 GCTTGGGAACCTGAGCGGTGGTAGCGGAAGATCAGACCATCTGATTCACCTTAACGGG 942  
Db 63 GCAAGGCCCGCGGAGGGGTGGGCGCGGAGGAGGTACAGATCCGCTTCCGCGCGCGG 122  
QY 943 TCTAGATTCTTTGGGC 958  
Db 123 GCGGGTGTGGGAC 138

## RESULT 11

US-08-311-731A-130  
; Sequence 130, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311,731A  
; FILING DATE: 530  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: C0044/7125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; INFORMATION FOR SEQ ID NO: 130:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36941 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: MYCOBACTERIUM LEPRAE  
US-08-311-731A-130

Query Match 3.3%; Score 31.8; DB 4; Length 36941;  
Best Local Similarity 54.8%; Pred. No. 26;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 454 CGAGCGGAAGCTCTCCAGGCTGGATCATCTCTGTCATCTCCCTGGTGGCGAGGAATTT 513  
Db 19602 CGGGAAGTAGTCTCCAGGCTGGCACTTGAGATCGAATTCGCGGTGCCGAGACCAA 19661  
QY 514 CTGCTCTCGCTCGCGGCGAGAAATCCATTCTCGAAGAAGATGTCGAAGAGTTTG 568  
Db 19662 CCGCACCGCACACGCGCTTTTGGCGGCGCCAAACAGTCTGGATGACCGCTCG 19716

## RESULT 12

US-08-506-296B-3/c  
; Sequence 3, Application US/08506296B  
; Patent No. 6313265  
; GENERAL INFORMATION:  
; APPLICANT: Phillips, Greg  
; APPLICANT: Cunningham, Bruce A.  
; APPLICANT: Crossin, Kathryn L.  
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES  
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 77  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute  
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: U.S.  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/506,296B  
; FILING DATE: 24-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: TSRI 488.0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 554-2937  
; TELEFAX: (619) 554-6312  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3991 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 59..3859  
US-08-506-296B-3

Query Match 3.3%; Score 31.6; DB 4; Length 3991;  
Best Local Similarity 47.9%; Pred. No. 8.1;  
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 391 ACGAAGGCTTCGAGGAATCTCTTTGTAGTCTTTCTTTGATGAAGCATCGAAGCGGTTAA 450  
Db 742 ACTGGGGGCCACGCGGAGGTCGAGGGGCTCTTCTGTGATGATGTCGCGGGGCGGAGGA 683  
QY 451 GCGGAGCGGAAGCTCTCCAGGCTGCGATGATCTGTGTGATCTCCTGTGCGGAGGA 510  
Db 682 GTGAGCGTGCAGATGATGTCGCGGGTGGCTGTCGCCACCATGGCGTTGAGAAGTAGAG 623  
QY 511 TTTCTCTCTCGCTCGCGGCGAGAAATCCATTCTCGAAGAAGATGTCGAAGAGTTTGTG 570  
Db 622 GTTCCCATCTGCCCCCATAGAGACCCCTCTCGTCTCGAGCGATGTGAACGATGTCGCTGT 563  
QY 571 CGCGGATTTG 580  
Db 562 GAGCCAATAG 553

## RESULT 13

US-09-375-975-39/c  
; Sequence 39, Application US/09375975  
; Patent No. 6593116  
; GENERAL INFORMATION:

APPLICANT: Huismann, Gjalte W.  
APPLICANT: Peoples, Oliver P.  
APPLICANT: Skraly, Frank A.  
TITLE OF INVENTION: Transgenic Microbial Polyhydroxyalkanoate Producers  
FILE REFERENCE: MBX 033  
CURRENT APPLICATION NUMBER: US/09/375,975  
CURRENT FILING DATE: 1999-08-17  
EARLIER APPLICATION NUMBER: 60/096,852  
EARLIER FILING DATE: 1998-09-18  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 39  
LENGTH: 1680  
TYPE: DNA  
ORGANISM: Pseudomonas oleovorans  
US-09-375-975-39

Query Match 3.2%; Score 31.2; DB 4; Length 1680;  
Best Local Similarity 44.2%; Pred. No. 6; Mismatches 163; Indels 0; Gaps 0;  
Matches 129; Conservative 0;

QY 181 GTTCAGAGATCATGACTGGTCCGGCGAGCATGGGCCAAGCCGTACGCCACCGGAAGT 240  
DB 504 GTTGACCAAGTCTTTGGCAGGTTGGACAGGCCATCGACAGGCTCTTGGCCGCGGTTTC 445  
QY 241 CGAGAAGATTCGGTAAGAAGAGCTTCGGCAGTGGTTGATAATGGCTTCGCCGCTACTT 300  
DB 444 GAAGAAGCGCTTGACTGCTGCGGGTTGGACAGAGGTTGGTGGCGAGCATGGCTTCGGT 385  
QY 301 RATCGGACAAATTTAAACATGCCACGAATCGCAGCTCAACGAGTACTGCTTGTCTGAC 360  
DB 384 CATCAGTTGATGACGAATCGCCGGCTGATGTCCTGGGGCGACAGTCTGCTGTGCC 325  
QY 361 GAAGCATTAACGCGCTAAGGCAATATGGAACGGAAGCTTCGAGGAATCTCTTGTAGTC 420  
DB 324 GATCCAGTCTCTCAGTCTCTGGCCAGCCAGATAGTTTGCAGGTAGCGCGGTAAG 265  
QY 421 TTCTTGATGAGCTGAGCGGTTAAAGCGGTAAGCGCGAGCGAGCTCTCCAGG 472  
DB 264 TGGGTTGTTGCTCCATGCGCGGTTCATTGAAGCGACGCTGCTGCTTTCCGGG 213

RESULT 14  
US-09-105-537-1/c  
Sequence 1, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 15872  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

Query Match 3.2%; Score 31.2; DB 3; Length 15872;  
Best Local Similarity 44.2%; Pred. No. 25; Mismatches 163; Indels 0; Gaps 0;  
Matches 129; Conservative 0;

QY 181 GTTCAGAGATCATGACTGGTCCGGCGAGCATGGGCCAAGCCGTACGCCACCGGAAGT 240  
DB 14651 GTTGACCAAGTCTTTGGCCAGGTTGGACAGGCCATCGACAGGCTCTTGGCCGCGGTTTC 14592  
QY 241 CGAGAAGATTCGTAAGAAGAGCGCTTCGCAAGTGGTTGATAATGGCTTCGCCGCTACTT 300

Query Match 3.2%; Score 31.2; DB 3; Length 15872;  
Best Local Similarity 44.2%; Pred. No. 25; Mismatches 163; Indels 0; Gaps 0;  
Matches 129; Conservative 0;

QY 181 GTTCAGAGATTCGTAAGAAGAGCGCTTCGCAAGTGGTTGATAATGGCTTCGCCGCTACTT 300  
DB 14651 GTTGACCAAGTCTTTGGCCAGGTTGGACAGGCCATCGACAGGCTCTTGGCCGCGGTTTC 14592  
QY 241 CGAGAAGATTCGTAAGAAGAGCGCTTCGCAAGTGGTTGATAATGGCTTCGCCGCTACTT 300

DB 14591 GAAGAAGCGTTTGACTGCTGCCGGTTGGACAGAGGTTGTGTCGAGCAATGGCTTCGGT 14532  
QY 301 AATCGGACAATTTAAACATCGCCAGCAATCGCAGCTCAACGAGTACTGCTTGTCTGAC 360  
DB 14531 CATCAGTTGATGAGCAACTGCGCGGCTGATGTCCTGGGGCGACAGTCTGCTGTGCC 14472  
QY 361 GAAGCATTTACAGGCGCTAAAGGCGATATGGAACGGAAGCTTCGAGGAATCTCTTGTAGTC 420  
DB 14471 GATCCAGTCTCTGAGCTCTTGGCCAGCGCCAGATAGGTTTGCAGGTAGCGCGGTAAG 14412  
QY 421 TTCTTGATGAAGCATCGAAGCGGTTAAAGCGGAGCGGAGCTCTCCAGG 472  
DB 14411 TGGGTTGTTGCTCCATGCGCGGTTCATTGAAGCGAGCGTCTGCTGCTTCCGGG 14360

## RESULT 15

US-09-091-609-1/c  
Sequence 1, Application US/09091609  
Patent No. 6600029  
GENERAL INFORMATION:  
APPLICANT: SHERMAN, DAVID H.  
APPLICANT: WILLIAMS, MARK D.  
APPLICANT: XUE, YONGQUAN  
TITLE OF INVENTION: METABOLIC ENGINEERING OF  
TITLE OF INVENTION: POLYHYDROXYALKANOATE MONOMER SYNTHASES  
FILE REFERENCE: 600.297US2  
CURRENT APPLICATION NUMBER: US/09/091,609  
CURRENT FILING DATE: 1998-06-19  
EARLIER APPLICATION NUMBER: PCT/US96/20119  
EARLIER FILING DATE: 1996-12-18  
EARLIER APPLICATION NUMBER: 60/008,847  
EARLIER FILING DATE: 1995-12-19  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 15872  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (20)...(13909)  
US-09-091-609-1

Query Match 3.2%; Score 31.2; DB 4; Length 15872;  
Best Local Similarity 44.2%; Pred. No. 25; Mismatches 163; Indels 0; Gaps 0;  
Matches 129; Conservative 0;

QY 181 GTTCAGAGATCATGACTGGTCCGGCGAGCATGGGCCAAGCCGTACGCCACCGGAAGT 240  
DB 14651 GTTGACCAAGTCTTTGGCCAGGTTGGACAGGCCATCGACAGGCTCTTGGCCGCGGTTTC 14592  
QY 241 CGAGAAGATTCGTAAGAAGAGCGCTTCGCGAGTGGTTGATAATGGCTTCGCCGCTACTT 300  
DB 14591 GAAGAAGCGTTTGTAGCTGCTGCGGGTTGACAGGGTTGGTGGAGACCATGGCTTCGGT 14532  
QY 301 AATCGGACAATTTAAACATCGCCAGCAACTCGCACTCGAGCTCAACGAGTACTGCTGTGAC 360  
DB 14531 CATCAGTTGATGACGAACCTGGCCGCGCTGATGTCCTGGGGCGACAGGTCGCTGTGCC 14472  
QY 361 GAAGCATTAACGCGCTAAGGCGATATGGAACGGAAGCTTCGAGGAATCTCTTGTAGTC 420  
DB 14471 GATCCAGTCTCTGAGCTCTTGGCCAGCGCCAGATAGGTTTGCAGGTAGCGCGGTAAG 14412  
QY 421 TTCTTGATGAAGCATCGAAGCGGTTAAAGCGGAGCGGAGCTCTCCAGG 472  
DB 14411 TGGGTTGTTGCTCCATGCGCGGTTCATTGAAGCGAGCGTCTGCTGCTTCCGGG 14360

Search completed: October 2, 2004, 05:16:08  
Job time : 83.5043 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2004, 03:13:49 ; Search time 504.178 Seconds  
(without alignments)  
9773.579 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972

Sequence: 1 gtaacagcagtgacgcat.....ttggccgtacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38.6	4.0	1163020	16	US-10-398-221-10
C 2	38.6	4.0	3011208	16	US-10-398-221-10
C 3	38.2	3.9	521	13	US-10-424-599-91751
C 4	37.8	3.9	802	15	US-10-184-644-312
C 5	37.8	3.9	802	15	US-10-184-634-312
C 6	36	3.7	2061	16	US-10-369-493-33642
C 7	35.8	3.7	78953	16	US-10-184-644-402
C 8	35.2	3.6	653	15	US-10-184-634-402
C 9	35.2	3.6	653	15	US-10-184-644-402
C 10	34.8	3.6	1590	14	US-10-041-472-4
C 11	34.4	3.5	1376	15	US-10-241-596-71
C 12	34.4	3.5	1376	15	US-10-241-596-73
C 13	34.4	3.5	1376	15	US-10-241-596-75
C 14	34.4	3.5	1376	15	US-10-241-596-77

C 15	34.4	3.5	1385	15	US-10-241-596-69	Sequence 69, Appl
C 16	34.4	3.5	2600	15	US-10-241-596-33	Sequence 33, Appl
C 17	34.4	3.5	2600	15	US-10-241-596-35	Sequence 35, Appl
C 18	34.4	3.5	2600	15	US-10-241-596-37	Sequence 37, Appl
C 19	34.4	3.5	2609	15	US-10-241-596-31	Sequence 31, Appl
C 20	34.4	3.5	2619	15	US-10-241-596-152	Sequence 152, App
C 21	34.4	3.5	2625	15	US-10-241-596-150	Sequence 150, App
C 22	34.4	3.5	2625	15	US-10-241-596-164	Sequence 164, App
C 23	34.4	3.5	2625	15	US-10-241-596-166	Sequence 166, App
C 24	34.4	3.5	2637	15	US-10-241-596-168	Sequence 168, App
C 25	34.4	3.5	2637	15	US-10-241-596-156	Sequence 156, App
C 26	34.4	3.5	2640	15	US-10-241-596-29	Sequence 29, Appl
C 27	34.4	3.5	2643	15	US-10-241-596-61	Sequence 61, Appl
C 28	34.4	3.5	2643	15	US-10-241-596-158	Sequence 158, App
C 29	34.4	3.5	2667	15	US-10-241-596-160	Sequence 160, App
C 30	34.4	3.5	3382	15	US-10-241-596-39	Sequence 39, Appl
C 31	34.4	3.5	3382	15	US-10-241-596-49	Sequence 49, Appl
C 32	34.4	3.5	3382	15	US-10-241-596-53	Sequence 53, Appl
C 33	34.4	3.5	3382	15	US-10-241-596-57	Sequence 57, Appl
C 34	34.4	3.5	3388	15	US-10-241-596-41	Sequence 41, Appl
C 35	34.4	3.5	3388	15	US-10-241-596-47	Sequence 47, Appl
C 36	34.4	3.5	3388	15	US-10-241-596-51	Sequence 51, Appl
C 37	34.4	3.5	3388	15	US-10-241-596-55	Sequence 55, Appl
C 38	34.4	3.5	3391	15	US-10-241-596-43	Sequence 43, Appl
C 39	34.4	3.5	3397	15	US-10-241-596-45	Sequence 45, Appl
C 40	34.2	3.5	448	13	US-10-424-599-77596	Sequence 77596, A
C 41	34.2	3.5	777	15	US-10-184-634-348	Sequence 348, App
C 42	34.2	3.5	777	15	US-10-184-634-348	Sequence 348, App
C 43	34	3.5	653	17	US-10-767-701-10779	Sequence 10779, A
C 44	33.8	3.5	4298	13	US-10-424-599-70962	Sequence 70962, A
C 45	33.8	3.5	7133	17	US-10-432-241-1	Sequence 1, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-398-221-10/c  
; Sequence 10, Application US/10398221  
; Publication No. US20040018514A1  
; GENERAL INFORMATION:  
; APPLICANT: KUNST, Frederik  
; APPLICANT: GLASER, Philippe  
; TITLE OF INVENTION: Listeria innocua, genome and applications  
; FILE REFERENCE: 344 702 - US  
; CURRENT APPLICATION NUMBER: US/10/398,221  
; CURRENT FILING DATE: 2003-03-27  
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: FR 00/12 697  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4025  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1163020  
; TYPE: DNA  
; ORGANISM: Listeria innocua  
; NAME/KEY: misc feature  
; LOCATION: (1)-(end)  
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u  
US-10-398-221-10

Query Match 4.0%; Score 38.6; DB 16; Length 1163020;  
Best Local Similarity 61.4%; Pred. No. 2;  
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 453 GGAGCGGAGCTCTCCAGGTGCGATGATCTCTGCGATCTCCCTGTGGCGAGGAAAT 512  
Db 513671 GCAGCGGCAAAATTTGGCAGCGCTCGATGAAATTTATCAAGTTGGCGGTGGCGACGCAATC 513612  
QY 513 TCTGCTCGCTCGCGGGCGGCAATCCATTGCTGAGAGAGA 553

Db 513611 GGTGCCCTAGCTTACGGTACAGCATCCATTCTTAAGTAGA 513571

## RESULT 2

```

US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/003 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
; US-10-398-221-2058

```

### RESULT 3

```

US-10-424-599-91751/c
; Sequence 91751, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91751
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MFT3847_53865C.1
US-10-424-599-91751

```

## RESULT 4

```

US-10-184-644-312/C
; Sequence 312, Application US/10184644
; Publication No. US2003004930A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C227
CURRENT APPLICATION NUMBER: US/10/184,644
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 312
LENGTH: 802
TYPE: PRT
ORGANISM: Homo Sapien
US-10-184-644-312

Query Match          3.9%; Score 37.8; DB 15; Length 802;
Best Local Similarity 4.8%; Pred. No. 0.1;
Matches 24; Conservative 169; Mismatches 312; Indels 0; Gaps 0;

QY   132 TTTGAAGCCGGAGATCGCTGGTGTGACTGCACCACTCCAGCCCTGGCGTTCAGAAGAT 191
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   587 K . B . MTMH . NSHN . . BB . MEN . N . . GAM . . NK . YB . KH . YCS . . . . . N . MMHM . . 528
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   192 CATGCAGCTGGTGCCGGAGCATTTGGGCCAACCGCTAGCCACCAGAAAGTCGAGAAGATT 251
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   527 MD . SNSC . . . . . B . MN . SNS . S . SCS . N . YMT . . TH . . D . B . . MS . A . . . . . RTDY 468
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   252 CCGTAAGAAGACGGCTTCGGCAGTGGTGTGATAATAGGTTCGCCGTACTTAATCGGCACAA 311
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   467 B . BAKMCMMDMR . . . H . R . TCRK . NBYYRCR . . . M . H . YCYAMTNIN . K . KB . KB . 408
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   312 TTAAACATCCCAAGCACTCGCAGCTCAACGAGTACTGCTGTCTGACGAAGCATTTACA 371
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   407 HH . T . Y . TSHKKY . . . CHFTSR . RMCA . . W . . . . . Y . Y . . YK . S . . C . . H . MD . . 348
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   372 GGCGCTAAGGCATATGAAACGGAAGGCTTCGAGAAATCTCTTGTAGTCTTTCTTGATGA 431
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   347 ANSRMSK . B . YD . . AYDY . S . MYS . KRSD . YR . R . . NMN . D . A . T . D . MSND . HY . R 288
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   432 AGCATCGAAGCGGTAAAGCGCGAGCGGAAGCTCTCCAGCTCGCATGATCTCTGTCCA 491
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   287 CMNN . RN . R . . YK . RWBG . BCT . YVBGYCB . BH . DAARK . Y . GMC . D . BKYY . B . YS 228
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   492 TCTCCCTGGTGGCGAGGAATTTCTGCTCTCCGCTCCGCGGCGAGAAATCCATTGCTGAAGAA 551
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   227 K . BC . CKKG . N . C . . . T . MCK . AAC . CA . C . . . . SSN . SY . N . YHCM . RBHHHTYK 168
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   552 GATGCTCGAAGAGTTGTGGCGCGAATNTGACCTCGCTCGACGGGTCTCTACCTCGGGA 611
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   167 W . RWH . RDHDKRM . KK . SVM . . . RSHHECY . BB . . . BSAHSC . SS . . BMC . . AMSWA 108
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   612 TACTCGTGAAGCATTCCTCTATTC 636
Db    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 5

```
US-10-184-634-312/c
; Sequence 312, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 312
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-312

Query Match      3.9%; Score 37.8; DB 15; Length 802;
Best Local Similarity 4.8%; Pred. No. 0.1;
Matches 24; Conservative 169; Mismatches 312; Indels 0; Gaps 0;

QY 132 TTTGAAGCCGGGAGATCGCTGGGTGACTCGACCATCCACGCTTGGCGTTTCAGAGAT 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 587 K..B.MTMH.NSHN...BB.MEN.N...CAM..NCK.YB.KH.YCS.....N.MHM.. 528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 192 CATGACTGTGTCGGGAGCATTTGGCCCAAGCGGTACGCCACCGGAAGTCGAGAATTT 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 527 MD.SMSC....B.MM.SNS..S.SCS..N.YMT...TH...D.B...MS.A...RTDYY 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 252 CGGTAGAGAGACGCTTCGGCAGTGGGTGTAATAGCTTCGCCGTACTTAATCGGACAA 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 467 B.BAKMMCMDDMR....H..R.TCRK.NBYRCR...M.H..YCYAMTNTN.K.KB.. 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 312 TTTAAACATCGCCAGCAATCGCAGCTCAAGAGTACTGTCTCTGACGAGCAATACA 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 HH.T.Y.TSHUKY...CHETS.RMCA...W.....Y.Y...YK.S...C...H.MD.. 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 372 GCGGCTAAGGGCATATGGAACGGAAGCTTCGAGGAATCTTTGTAGTCTTTCTTGATGA 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 ANSRMSK.B..YD...AYDY..S.MYS.KRSD.YR.R..NMN.D.A.T.D.MSND.HY.R 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 432 AGCATGGAAGCGGTTAAAGCGGACGGAAGCTCTCCAGGTGCGATATCTCTCGA 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 CMWN..RN.R...YK.RSWBG.BCT.YYBGYCB.BH.DAARK.Y.GMC.D.BKYY.B.YS 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 492 TCTCCTGTGTGCGAGGAATTTCTGCTCGCTGCGGCGGAGAAATCCATCTCTGAAGA 551
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 K.BC.CCKG.N.C...T.MCK..AAC.CA.C.....SSN.SY..N..YHCM.RBHHYTK 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 552 GATGTCGAAGATTTGTCCCGGATTTGCACCTCGCTCGAGGCTGCTACTCTCGGGA 611
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 167 .W.RMH.RDHDKRM.KK.SMM...RSHHBCY..BB...BSAHSK.SS...BMC..AMSWA 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 612 TACTCGTGAAGACATTCCTTATTC 636
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 107 YRTTBTYA..RM.TAKBCS...NYH 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-10-369-493-33642
; Sequence 33642, Application US/10369493
; Publication No. US20030233675A1

GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33642
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-33642

Query Match      3.7%; Score 36; DB 16; Length 2061;
Best Local Similarity 55.6%; Pred. No. 0.66;
Matches 69; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 300 TAATCGGACAAATTTAAACATCGCCAGACTCGCAGCTCAACGAGTACTGCTTGTGA 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 612 TACGATGGCGCATTAGCCCTCGCCATACACAACAGCTGTTTGCCCAACGCGTGGGTGA 671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 360 CGAAGCATTACAGCGCGCTAAGGCGCATATGGAACGGAAGGCTTCGAGGAATCTCTGTAGT 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 672 TGAAGCCTTTTGGCCAAGTGGGCCAATGCAATGAGGCTTTAAGGCTATGTTAGGGC 731
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 420 CTTT 423
   |||
Db 732 GTTT 735
   |||

RESULT 7
US-10-085-117-31/c
; Sequence 31, Application US/10085117
; Publication No. US2003023334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 78953
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(78953)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-31

Query Match      3.7%; Score 35.8; DB 16; Length 78953;
Best Local Similarity 49.2%; Pred. No. 4.7;
Matches 94; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 596 TGCTCTACTCGGGGATCTCGTGAAGACCATTCCTATTTCGAACGAGAGATCTTTGAG 655
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10142 TGCACCACTAGGCGCTCAGTTGAAGATGTACACCCAGTCACTGAAGCGGTGGTAGCTG 10083
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 656 AGGTGCTCGGCTGACTTTTCGACCCCATGGTTCGAATGTCGAGACCTTATTCCTCATGAG 715
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10082 TTGACCTCGGCGGTCTCTCGCTTGGGTCACACTAGCTTGTCTGCTGCTGTTCGAA 10023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



QY 716 AAGTTCGGTGGCTTTCTTATGAGCGCGTGAAGGTAAAGGTAAAGGTCCGTTTCATGAGG 775  
Db 10022 ATGTAATTCACAGCTTTGCTGTGGGAAGTCCAGCAAGAAAGAGGTACCTTTATGGGG 9963  
QY 776 AGCGGCATCGC 786  
Db 9962 AGCAAGCAC 9952

RESULT 8  
US-10-184-644-402  
; Sequence 402, Application US/10184644  
; Publication No. US20030044930A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Goddard, Paul J.  
; APPLICANT: Goddard, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C227  
; CURRENT APPLICATION NUMBER: US/10/184,644  
; CURRENT FILING DATE: 2002-06-28  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 402  
; LENGTH: 653  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-184-644-402

Query Match 3.6%; Score 35.2; DB 15; Length 653;  
Best Local Similarity 6.4%; Pred. No. 0.7;  
Matches 28; Conservative 141; Mismatches 266; Indels 0; Gaps 0;  
QY 11 GTGACGGCATCGACGGAACGGTAGCGACATCGACTACTGCGGGCGGTCTGCTAAAGCGTT 70  
Db 88 SEINEDSDHYAIMPPLQFMEIPSMRRELFFRDIERGDIVIGRISIREFGFFFWVLIC 147  
QY 71 TTGGGTTTACGCGCAACGATATAACGTCGCTAGCGCTGTGACATTCGCTGCGCTTGGCG 130  
Db 148 LGSGIMRDIAHLEITALEPLRDPVSHNHDPLSYQTGDIIRAGIKIDIRYHEKLAIVSL 207  
QY 131 GTTGAAGCGGAGATCGCTGGTGTGACTCGACCATCCAGCCCTTGGCGTTCAGAAGA 190  
Db 208 YSSSLPPLHSGIKLVISSEELPLYRRSVLSNSLESYENVMQSLGFVNPVGFLL 267  
QY 191 TCATGACTGTCGCGGAGCATTCGGCCAGCGTACGCCACCGAAGTCGAGAAGATT 250  
Db 268 EKLGDENPPLSMRGLQSNFSEDDFASALRKQASWALKCVKIGVDYFKVGRHVDAM 327  
QY 251 TCCGTAAGAAGACGCTTCGGCAGTGGTGTGATAATGGCTTCGCGCTTAAATGCGGACA 310  
Db 328 NEYNKALEIDKQNVREALVARGALYATKSLNKAIEDFELALENCPTHRNARKYLCQTLVE 387  
QY 311 ATTTAAACATCCCGAGCAACTCGCAGCTCAACGAGTACTGCTGTCTGACGAAGCATTAC 370  
Db 388 RGGQLEEBEKFNLAEYSYKKALALDETFFKDAEDALQKLHKYMQKSLREKQAEKEKQK 447  
QY 371 AGCGCTTAAGGCATATGAACGAGGCTTCAGAAATCTCTGTAGCTTTCTTTGATG 430  
Db 448 TKIETSAREKRLKLEERLKKRKRKSTSSSSVSSADESVSSSSSSSGHKKRHKHR 507  
QY 431 AAGCATCGAAGCGG 445

RESULT 10  
US-10-041-472-4/c  
; Sequence 4, Application US/10041472  
; Publication No. US20020092039A1

Db 508 NRSESSRRRRSSR 522  
RESULT 9  
US-10-184-634-402  
; Sequence 402, Application US/10184634  
; Publication No. US200300684A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3430R1C217  
; CURRENT APPLICATION NUMBER: US/10/184,634  
; CURRENT FILING DATE: 2002-06-28  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 402  
; LENGTH: 653  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-184-634-402

Query Match 3.6%; Score 35.2; DB 15; Length 653;  
Best Local Similarity 6.4%; Pred. No. 0.7;  
Matches 28; Conservative 141; Mismatches 266; Indels 0; Gaps 0;  
QY 11 GTGACGGCATCGACGGAACGGTAGCGACATCGACTACTGCGGGCGGTCTGCTAAAGCGTT 70  
Db 88 SEINEDSDHYAIMPPLQFMEIPSMRRELFFRDIERGDIVIGRISIREFGFFFWVLIC 147  
QY 71 TTGGGTTTACGCGCAACGATATAACGTCGCTAGCGCTGTGACATTCGCTGCGCTTGGCG 130  
Db 148 LGSGIMRDIAHLEITALEPLRDPVSHNHDPLSYQTGDIIRAGIKIDIRYHEKLAIVSL 207  
QY 131 GTTGAAGCGGAGATCGCTGGTGTGACTCGACCATCCAGCCCTTGGCGTTCAGAAGA 190  
Db 208 YSSSLPPLHSGIKLVISSEELPLYRRSVLSNSLESYENVMQSLGFVNPVGFLL 267  
QY 191 TCATGACTGTCGCGGAGCATTCGGCCAGCGTACGCCACCGAAGTCGAGAAGATT 250  
Db 268 EKLGDENPPLSMRGLQSNFSEDDFASALRKQASWALKCVKIGVDYFKVGRHVDAM 327  
QY 251 TCCGTAAGAAGACGCTTCGGCAGTGGTGTGATAATGGCTTCGCGCTTAAATGCGGACA 310  
Db 328 NEYNKALEIDKQNVREALVARGALYATKSLNKAIEDFELALENCPTHRNARKYLCQTLVE 387  
QY 311 ATTTAAACATCCCGAGCAACTCGCAGCTCAACGAGTACTGCTGTCTGACGAAGCATTAC 370  
Db 388 RGGQLEEBEKFNLAEYSYKKALALDETFFKDAEDALQKLHKYMQKSLREKQAEKEKQK 447  
QY 371 AGCGCTTAAGGCATATGAACGAGGCTTCAGAAATCTCTGTAGCTTTCTTTGATG 430  
Db 448 TKIETSAREKRLKLEERLKKRKRKSTSSSSVSSADESVSSSSSSSGHKKRHKHR 507  
QY 431 AAGCATCGAAGCGG 445  
Db 508 NRSESSRRRRSSR 522

RESULT 10  
US-10-041-472-4/c  
; Sequence 4, Application US/10041472  
; Publication No. US20020092039A1

```
; GENERAL INFORMATION:
; APPLICANT: Shewmaker, Christine
; TITLE OF INVENTION: METHODS FOR PRODUCING CAROTENOID COMPOUNDS AND SPECIALITY OILS IN SEEDS
; TITLE OF INVENTION: SEEDS
; FILE REFERENCE: 16516.141
; CURRENT APPLICATION NUMBER: US/10/041.472
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 08/908758
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/024145
; PRIOR FILING DATE: 1996-09-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1590
; TYPE: DNA
; ORGANISM: Brassica napus
US-10-041-472-4

Query Match      3.6%; Score 34.8; DB 14; Length 1590;
Best Local Similarity 52.0%; Pred. No. 1.5;
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

QY 112 ACATTGCTCGCTTGGCGTTTGAACCGGGAGATCGCTGGGTGACTCGACCACTCCA 171
Db 1180 ACGGAGTTGGGTGACGAGTTCGAAGCCATGGACTGTGACTGCTCGACACCACT 1121

QY 172 CG--CCTTGGCGTTTCAGAACATCATGACTGGTCCGCGGAGCATTTGGGCCAAGCCGTACG 229
Db 1120 GGTCCGCGCGCTGCTGTCTACATCAACACGCGCTCCGAAAGGAGCCTGTCGCCGCTTACG 1061

QY 230 CCACCGGAAGTCGAGAGATTTCCGTAAGAAGACGCTTCGCGCAGTGGGTGTGATAATGGCT 289
Db 1060 GGAGAGTGAACCGTAAACAGCTCAATCCAGATGCTTTCAGAGTGATCACCACCGTG 1001

QY 290 TCGCCGTACTTAATGC 305
Db 1000 TTAGATTCCATCAGGC 985

RESULT 11
US-10-241-596-71/c
; Sequence 71, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.013003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-71

Query Match      3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 697 GACCTTATTCTCCATGACGAGTTCTGGTGGTGGCTTTTCTTATGGAGCGCGTGAAGAAGT 756
Db 1263 GAAGTTTTTTCAGTTTTTGTGAAGTTCATGTTGTGATTTCCGTGTTCTGGCGGTTAAAT 1204

QY 757 AAAGTCCGTTTGATGAGGAGCGCATCGAGCCTCGAAGAGCTATTTCGTTACACCTTCA 816
Db 1203 AGCAGCCAGGTGGTGTACGAGGTGAACCATCGTAGATAGTAGTTAACTTCGG 1144

QY 817 GCGGGTCTAAATTTTGTAAACTGCTTTG 844
Db 1143 CACGATGTGATTTTGAATACTGCTTTG 1116

RESULT 12
US-10-241-596-73/c
; Sequence 73, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.013003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-73

Query Match      3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 697 GACCTTATTCTCCATGACGAGTTCTGGTGGTGGCTTTTCTTATGGAGCGCGTGAAGAAGT 756
Db 1263 GAAGTTTTTTCAGTTTTTGTGAAGTTCATGTTGTGATTTCCGTGTTCTGGCGGTTAAAT 1204

QY 757 AAAGTCCGTTTGATGAGGAGCGCATCGAGCCTCGAAGAGCTATTTCGTTACACCTTCA 816
Db 1203 AGCAGCCAGGTGGTGTACGAGGTGAACCATCGTAGATAGTAGTTAACTTCGG 1144

QY 817 GCGGGTCTAAATTTTGTAAACTGCTTTG 844
Db 1143 CACGATGTGATTTTGAATACTGCTTTG 1116

RESULT 13
US-10-241-596-75/c
; Sequence 75, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.013003
; CURRENT APPLICATION NUMBER: US/10/241,596
```

```
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 697 GACCTTATTCTCCATGACGAGTTCTGGTGGTGGCTTTTCTTATGGAGCGCGTGAAGAAGT 756
Db 1263 GAAGTTTTTTCAGTTTTTGTGAAGTTCATGTTGTGATTTCCGTGTTCTGGCGGTTAAAT 1204

QY 757 AAAGTCCGTTTGATGAGGAGCGCATCGAGCCTCGAAGAGCTATTTCGTTACACCTTCA 816
Db 1203 AGCAGCCAGGTGGTGTACGAGGTGAACCATCGTAGATAGTAGTTAACTTCGG 1144

QY 817 GCGGGTCTAAATTTTGTAAACTGCTTTG 844
Db 1143 CACGATGTGATTTTGAATACTGCTTTG 1116
```

```
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-75

Query Match          3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 697 GACCTTATTCCTCCATGACGAGTTCGTGGTGGCTTTTCTTATGAGGCGGTGAAAAGT 756
Db 1263 GAAGTTTTTTCAGTTTGTGGAAGTTCATGTTGTTGATTTCCGTGTTCTGCGCGTTAAAAAT 1204

Qy 757 AAAGTCCGTTTGATGAGGCGGCATCGCAGCTTCAAGAGCTATTTCGTTACACCTTCA 816
Db 1203 AGCAGCCAGGTGGTGTACGCAAGTTGAAACCATCGTAGATAGTGTAGTTAACTTCGG 1144

Qy 817 GCGGCTCTAAATTTTGTAAACTGCTTTG 844
Db 1143 CACGATGTTGATTTTGAATCTGCTTG 1116

RESULT 15
US-10-241-596-69/c
; Sequence 69, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-69

Query Match          3.5%; Score 34.4; DB 15; Length 1385;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 697 GACCTTATTCCTCCATGACGAGTTCGTGGTGGCTTTTCTTATGAGGCGGTGAAAAGT 756
Db 1263 GAAGTTTTTTCAGTTTGTGGAAGTTCATGTTGTTGATTTCCGTGTTCTGCGCGTTAAAAAT 1204

Qy 757 AAAGTCCGTTTGATGAGGCGGCATCGCAGCTTCAAGAGCTATTTCGTTACACCTTCA 816
Db 1203 AGCAGCCAGGTGGTGTACGCAAGTTGAAACCATCGTAGATAGTGTAGTTAACTTCGG 1144

Qy 817 GCGGCTCTAAATTTTGTAAACTGCTTTG 844
Db 1143 CACGATGTTGATTTTGAATCTGCTTG 1116

Search completed: October 2, 2004, 08:33:28
Job time : 510.178 secs
```

```
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-75

Query Match          3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 697 GACCTTATTCCTCCATGACGAGTTCGTGGTGGCTTTTCTTATGAGGCGGTGAAAAGT 756
Db 1263 GAAGTTTTTTCAGTTTGTGGAAGTTCATGTTGTTGATTTCCGTGTTCTGCGCGTTAAAAAT 1204

Qy 757 AAAGTCCGTTTGATGAGGCGGCATCGCAGCTTCAAGAGCTATTTCGTTACACCTTCA 816
Db 1203 AGCAGCCAGGTGGTGTACGCAAGTTGAAACCATCGTAGATAGTGTAGTTAACTTCGG 1144

Qy 817 GCGGCTCTAAATTTTGTAAACTGCTTTG 844
Db 1143 CACGATGTTGATTTTGAATCTGCTTG 1116

RESULT 14
US-10-241-596-77/c
; Sequence 77, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-77

Query Match          3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 11.1792 Seconds  
(without alignments)  
2779.248 Million cell updates/sec

Title: US-10-668-047-4  
Perfect score: 1689  
Sequence: 1 VNSSDGIDGTVASIDTARAL.....EDPDHLHLNGSRFLGFIYER 323

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	632.5	37.4	546	2 AF2268	type II site-speci
2	603.5	35.7	316	2 S35516	type II site-speci
3	511	30.3	326	1 NDORF	type II site-speci
4	104	6.2	772	1 JQ2024	outer layer protei
5	100	5.9	772	1 JQ2025	outer layer protei
6	100	5.9	772	2 A46110	outer capsid prote
7	99.5	5.9	1203	2 H87687	outer capsid prote
8	97.5	5.8	289	2 B83189	helicase, UvrD/Rep
9	97	5.7	485	2 E83663	elongation factor
10	96	5.7	516	2 H72427	glutamyl-tRNA synt
11	95.5	5.7	797	2 G83570	DNA mismatch repai
12	95.5	5.7	1057	2 F89892	probable two-compo
13	95	5.6	409	2 AD3118	carbamoyl-phosphat
14	95	5.6	409	2 C98169	ABC transporter, s
15	95	5.6	4574	2 G02520	hypothetical prote
16	95	5.6	4684	2 A59404	plectin - human
17	94.5	5.6	574	2 T08434	plectin [imported]
18	94.5	5.6	619	2 A69352	misato protein - f
19	94	5.6	264	2 AD3353	conserved hypothet
20	94	5.6	312	2 S76937	probable membrane
21	94	5.6	772	1 A44052	hypothetical prote
22	93.5	5.5	359	2 A46038	outer layer protei
23	93.5	5.5	399	2 B69069	probable transcrip
24	93	5.5	772	2 A46108	GTP-binding protei
25	92.5	5.5	449	2 D75033	outer capsid prote
26	92	5.4	3951	1 VF1HB1	probable prolifera
27	91.5	5.4	318	2 F84495	F1 protein - avian
28	91.5	5.4	788	1 JCVLHH	probable Athila re
29	91.5	5.4	1360	2 JCV839	DNA-directed DNA p
					GTBP-N protein - h

30	90.5	5.4	406	2 H70364	iron-sulfur cofact
31	90.5	5.4	456	2 F47677	phosphomannomutase
32	90	5.3	474	2 T19219	hypothetical prote
33	90	5.3	480	1 N1A1MA	nitrogenase (EC 1.
34	90	5.3	497	2 AF1986	nitrogenase molybd
35	90	5.3	852	2 T28790	hypothetical prote
36	89.5	5.3	360	2 A75315	probable chalcone
37	89.5	5.3	1010	2 T36383	probable large ATP
38	89	5.3	666	2 S75289	ribonuclease II -
39	88.5	5.2	489	1 N1BCAT	nitrogenase (EC 1.
40	88.5	5.2	504	2 S05197	signal recognition
41	88.5	5.2	504	2 S54143	SRP 54 protein - h
42	88	5.2	367	2 E83676	pyruvate dehydroge
43	88	5.2	408	2 H90222	GTP binding protei
44	88	5.2	772	1 JQ2026	outer layer protei
45	88	5.2	1965	2 S75200	fat protein - Syne

ALIGNMENTS

RESULT 1

AF2268  
type II site-specific deoxyribonuclease [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AF2268  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF2268  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-546 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA075400.1; PID:G17132835; GSPDB:GNO0179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr3701

Query Match	37.4%;	Score	632.5;	DB	2;	Length	546;
Best Local Similarity	40.2%;	Pred. No.	1.5e-42;				
Matches	129;	Conservative	65;	Mismatches	122;	Indels	5;
Gaps	3;						
QY	4	SDGIDGTVAS--IDTARALLKRFGFDQRYNRSVATLLALAGLKPGRWVDSTTPRLGV	61				
Db	223	SEAIKAVLAAKRIEALAILKDISAPKEQONERSALFLALADIRPEIPWQTATSPRRRI	282				
QY	62	OKIMDSGHEHWAQPYATGSRDPKKTILQWVDNGFAVLNADMLNATNSQLNEYCLISDE	121				
Db	283	TEMMDWFRDHYGQYAPNTRVRRQTMHQFVQMGIWVENPDQDFRINSKPKWCYQLHQQ	342				
QY	122	ALQALRAYTEGFEGLSVFLDEASKAVKARAEALQAAISVDLPGGEFFLSPAGONPL	181				
Db	343	FVTLKSYSGEWEETRRNIVSKMLLQDRNRI--PMIPVSLPNGQAIQISSGGONIL	400				
QY	182	LKKMVEEFVPRFAPRSTVLYLGDTRGKLSLFFEREIFEVVLGLTFDPRGMPDLILHDEV	241				
Db	401	IKEILNFCPRFTPEGLVLFVGDAGNKFVNQTKFREI-GIELDPHGKMDIVVYERQ	459				
QY	242	GMLFLMEAVKSGPPDEEHRSLQELFVTPSAGLIPVNCFNENRSMQWLPCLAMTEAW	301				
Db	460	EWLVLIEAVTSGPVMNLKRNELKRLFQSSRQGLFVFTAFPSRKEMTRYLAETSWEDEV	519				
QY	302	VAEDPDHLHLNGSRFLGPEY	322				
Db	520	VAAQPDHMIHNGERFLGPEY	540				
RESULT 2							
S35516							



259 ERHRSLOELFVTPSAGLIFVNCPENRESMROWLP 294  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
733 TREQALNMLRSDPKVLRSFIN--QNNPIKNRIEQL 766

RESULT 5  
JQ2025  
outer layer protein VP4 - bovine rotavirus B (serotype 223)  
N:Alternate names: hemagglutinin; outer capsid protein VP4  
C:Contains: outer layer protein VP5; outer layer protein VP8  
C:Species: bovine rotavirus B  
C:Date: 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 02-Jun-2000  
C:Accession: JQ2025  
J:Taniguchi, K.; Urasawa, T.; Urasawa, S.  
R:Gen. Virol. 74, 1215-1221, 1993  
A:Title: Independent segregation of the VP4 and the VP7 genes in bovine rotaviruses as  
A:Reference number: JQ2022; MUID:93286580; PMID:8389807  
A:Accession: JQ2025  
A:Molecule type: genomic RNA  
A:Residues: 1-772 <TAN>  
A:Cross-references: GB:DL3394; NID:G939328  
A:Note: this translation is not annotated in GenBank entry ROBB223VP4, release 113.0  
C:Genetics:  
A:Map position: segment 4  
C:Superfamily: rotavirus outer layer protein VP3  
C:Keywords: glycoprotein; hemagglutinin; outer capsid protein  
F:1-242/Product: outer layer protein VP8 #status predicted <VP8>  
F:243-248/Region: cleavage processing #status predicted  
F:249-772/Product: outer layer protein VP5 #status predicted <VP5>  
F:32,98,116,123,127,173/Binding site: carbohydrate (Asn) #status predicted

Query Match 5.9%; Score 100; DB 1; Length 772;  
Best Local Similarity 22.9%; Pred. No. 7.2;  
Matches 77; Conservative 46; Mismatches 129; Indels 84; Gaps 16;

Qy 1 VNSDGDIGTGVASIDTARALLKRGFGDAQYNVRSVATLLALA-----GLKPGDRWVDST 55  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
473 VPSNDDYQTPANSVTVRQDLER-QLDEMEREFEINLSANIALSQLIDLALLP---LDMF 527

Qy 56 TPLRGVQKIMDSGEHWAKPYATGSRDEPK-----KTIHQ 91  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
528 SMFSGIQTVE-----AAKFTATSVMKFRKSLAKSVNSLTDAITDAAGISRSSTLS 582

Qy 92 -----WVDNGFAVLNADNLNIATNSQLNEYCLSDREALQALRAYGTGEF--EESLVVFL 142  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
583 VNSAASVMTDISDIVDSTDNVAAAT-----ATAAKKFRVKEFTTEFNGVSFD 630

Qy 143 DEASKAVKARAEALQAAMISVDLPGGEEFLLSPAGQNPLKKWVEFVFPAPR---STV 199  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
631 DISAAVVKTKMSKLV-----VD-----PQIITEASEKFI PNRYRLIDGEX 676

Qy 200 LYLGDTRGRKHSLEFERBIFEVGLGTFPHGRMPDLILHDEVGRWLMEAVKS-KGPFDE 258  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
677 VYEVTTGKYFAVLTETFEVV--FDAE-RFAELVTDSPVISAIDFTIKNLNDNYGI 732

Qy 259 ERHRSLOELFVTPSAGLIFVNCPENRESMROWLP 294  
Db |::||| |::||| |::||| |::||| |::||| |::|||  
733 TREQALNMLRSDPKVLRSFIN--QNNPIKNRIEQL 766

RESULT 6  
A46110  
outer capsid protein VP4 - human rotavirus (strain 116E)  
C:Species: human rotavirus  
C:Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Sep-1999  
C:Accession: A46110  
R:Gentsch, J.R.; Das, B.K.; Jiang, B.; Bhan, M.K.; Glass, R.I.  
Virology 194, 424-430, 1993  
A:Title: Similarity of the VP4 protein of human rotavirus strain 116E to that of the bo  
A:Reference number: A46110; MUID:93242780; PMID:8386888  
A:Contents: 116E  
A:Accession: A46110





```

QY 275 LIFVNC-----PENRESM-----ROWLPDLAWE---TEA 300
      ::|||
Db 287 VLFIEVNPEDVNIHPSKIIVKFLNEBKVKKSLEETLKRNLARKWRSVAYEISRRAL 346
      |||||
QY 301 WVAEPDGH 308
      |||||
Db 347 SVAAEPESH 354

RESULT 11
G83570
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83570
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoquchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83570
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <STO>
A:Cross-references: GB:AE004496; GB:AE004091; NID:G9946468; PIDN:AAG03989.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0600

Query Match 5.7%; Score 95.5; DB 2; Length 797;
Best Local Similarity 24.7%; Pred. No. 17;
Matches 80; Conservative 36; Mismatches 91; Indels 117; Gaps 22;

QY 10 TVASIDTARALKRFGFDAQ-RYNVRSVAVILLA-----LAGLKPQ---DRWVDST 55
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 325 TQFSIDNSTGLIWNVDKRVRYANRAAEQMLGHADGQLVDRPLADFPGLMDRWLNL- 383
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 56 TPLRGVQKIMDSGEHWAKPYATGSRSE---DFRKKTLRQWVDNGFAVLNADNLNATNSQ 112
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 384 -----WRR--ARNSEGLPSFETRCILR-----AD----- 405
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 113 LNEYCLSDPALQALRAYGTGFPESLVFLDEASKAVKARAEALQ---AAM--ISVDLPQ 167
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 406 -GSWLPADVSLSLFR-FGT---SEYLVVFLSDVTRRRAR-EALQESARKMGKIASNVPG 459
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 168 GEEFLLSPAGQNPLLKKVVEEFV-PRPAPRSTVYLIGDTRGKHSIF---EREIPEEVIGL 223
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 460 -----MVFRLRPRPAGAFSDFAYISE--GSEALVGVSAARELIESGRGI 500
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 224 TFDPHGRMPDLILHDEVRGMLFMEAVKSGPFDDEERHSLSLOELFVTPSAGLIFVN---- 279
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 501 RGLVH---PD-----DRERTWSSQMAAL-----DENRDWHQGRILTRQGLRWADIKAS 547
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 280 --CFENRESMROWLPDLAWETEAW 301
      ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 548 ARCFEDGRAV-----WDGVVW 563

RESULT 12
F89892
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F89892
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogudma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89892
A>Status: preliminary

```



QY	137	-----SIIVFLDASAKVAKARAEALQAMTISVLPGGEELLSPAGQNPQLIKKWVEFVP	191
DB	2254	EEELFSVQMEBLSK-LKARTEAENRALLRDKNTQRLQEEAEK---MKQVAEE---	2306
QY	192	RFAPRSTVLYLGTGRKHSLFEREIEFVVLGLTFDPHGMRPDLLIHDEVRGMFLMEAVK	251
DB	2307	--AARLSV-----AAQEAARLSQABEDLA-----QORAAEKWLKKMQA---VQEAIR	2351
QY	252	SKGPFD-----BERHRSLOELFVTPSAGLIFVNCFNRRSMRQWLPFLAWETAW	301
DB	2352	LKAEAEILQOQKEIAEQAARLQ-----EDKQMAQ---QLAEETQGF	2391

Search completed: October 1, 2004, 16:11:04  
Job time : 15.1792 secs

```

179  --NPLLKKQWVEEV-----PRFAPRSTVL 200
      | | | | |
      | | | | |
364  VTNDLMPKAVQEHVLPQSVYPRLSLPAFWMDSAL 398
      | | | | |
      | | | | |

RESULT 14
C981169
Hypothetical protein AGR_L_608 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: C981169
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2322-2328, 2001
A;Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: C981169
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-409 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88877.1; PID:g15158644; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_608
A;Map position: linear chromosome

```

Query Match	5.6%	Score 95	DB 2	Length 409
Best Local Similarity	26.0%	Pred. No. 7.4		
Matches 56	Conservative 23	Mismatches 62	Indels 74	Gaps 12
Qy	51	WVDTT <sup>     </sup> RLRGVQ-KINDWSGEHM-----AKPVATGSR <sup>     </sup> EDFRK-----KTLRQ-----	91	
Db	193	WAEKNPK <sup>     </sup> RLGVTLPKAGSGGGFYISVALNVLIG---DCRKQLTDY <sup>     </sup> NQTLQQAEDWAMES	249	
Qy	92	-----W-----VNGPFAVLNADNINIATNSQLNEYCLSDBALQALRAYGTG <sup>     </sup> EGF	134	
Db	250	ECLTPVWDY <sup>     </sup> YRRLLA <sup>     </sup> VAELTNG-----NADTLINLNQO <sup>     </sup> LYMTGV <sup>     </sup> WED--QVMSFLGT <sup>     </sup> KQL	303	
Qy	135	EESLAV <sup>     </sup> VFLEASKA-----VKARAAALQAA <sup>     </sup> MSVDELPG <sup>     </sup> GE <sup>     </sup> RFLL-----SPAQ	178	
Db	304	PDSE <sup>     </sup> RVILLKE <sup>     </sup> QGVSGDAMFVPANAKNVASALLI <sup>     </sup> DMMSKE <sup>     </sup> FLMKLENKAS <sup>     </sup> RSRPTD	363	
Qy	179	--NPLLK <sup>     </sup> KKWDE <sup>     </sup> EFV-----PRPAP <sup>     </sup> STVL	200	
Db	364	VTNDLMPK <sup>     </sup> AVOE <sup>     </sup> HVLPQSVY <sup>     </sup> RLSLP <sup>     </sup> AFDM <sup>     </sup> STAL	398	

RESULT 15  
G02520  
plectin - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
C:Accession: G02520  
R:McLean, W.H.I.; Smith, F.J.D.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: H01385  
A:Accession: G02520  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4574 <MCL>  
A:Cross-references: EMBL:U53204; NID:G1477645; PIDN:AA05427.1; PID:G1477646  
C:Genetics:  
A:Gene: PLEC1  
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S  
P:68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

```
Query Match      5.6%; Score 95; DB 2; Length 4574;
Best Local Similarity 25.9%; Pred.No. 2.2e+02;
Matches 62; Conservative 39; Mismatches 70; Indels 68; Gaps 14;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 7.58591 Seconds  
(without alignments)  
2217.092 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIDGTVASIDTARALL.....EDPDHLHLNGSRFLGPFYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	603.5	35.7	316	1 T2BB_BACSU	P33562 bacillus su
2	511	30.3	326	1 T2PS_PROST	P00640 providencia
3	104	6.2	772	1 VP4_ROTBA	P36307 bovine rota
4	101.5	6.0	1057	1 CARB_STAEP	O8C914 staphylococ
5	101	6.0	772	1 VP4_ROTBB	P35746 bovine rota
6	100	6.0	772	1 VP4_ROTBU	O09113 human rotav
7	97.5	5.8	289	1 EFTS_PSEAE	O82851 pseudomonas
8	97	5.7	485	1 SYE_BACHD	O9kgf6 bacillus ha
9	96.5	5.7	513	1 C942_VICSA	P74925 thermotoga
10	96	5.7	510	1 MUTL_THEMA	O99ur5 staphylococ
11	95.5	5.7	1057	1 CARB_STAAM	P58940 staphylococ
12	95.5	5.7	1057	1 CARB_STAAM	Q15149 homo sapien
13	95	5.6	4684	1 PLEI_HUMAN	O896d5 bradyrhizob
14	94.5	5.6	454	1 TOLB_BRAJA	O29441 archaeoglob
15	94.5	5.6	619	1 Y181_ARCFU	O9y4k1 homo sapien
16	93	5.5	1723	1 A1M1_HUMAN	O8fp92 corynebacte
17	92.5	5.5	492	1 Y193_COREF	O83z25 enterococcu
18	92	5.4	447	1 DNAA_ENTEA	P27920 a replicase
19	92	5.4	6629	1 RIAB_IBVB	O91qt2 a replicase
20	92	5.4	6629	1 RIAB_IBVCB	P13846 heron hepat
21	91.5	5.4	788	1 DPOL_HPBVC	P52701 homo sapien
22	91.5	5.4	1360	1 MSHE_HUMAN	O01411 salmonella
23	90.5	5.4	456	1 MANB_SALMO	P00464 anabaena sp
24	90	5.3	497	1 N1FD_ANASP	Q98h64 rhizobium l
25	89.5	5.3	485	1 HSS_RHILQ	P73177 synchocyst
26	89	5.3	666	1 RNZH_SNY3	P06662 thiobacillu
27	88.5	5.2	489	1 N1FD_THIFE	P13624 homo sapien
28	88.5	5.2	504	1 SR54_HUMAN	P57906 pasteurella
29	88	5.2	480	1 SYE_FASMU	O8df94 vibrio vuln
30	88	5.2	553	1 SNTD_VIBVU	P36308 bovine rota
31	88	5.2	772	1 VP4_ROTBU	O00763 homo sapien
32	88	5.2	2483	1 COA2_HUMAN	P14576 mus musculu
33	87.5	5.2	504	1 SR54_MOUSE	

```

34 87.5 5.2 701 1 PALL_ORYSA P14717 oryza sativ
35 87.5 5.2 872 1 SYA_LACIA Q9cew0 lactococcus
36 87 5.2 466 1 SYC_BACHD Q9kgf4 bacillus ha
37 87 5.2 737 1 RECG_MYCTU P95122 mycobacteri
38 87 5.2 1215 1 ATC6_YEAST P39986 saccharomyc
39 86.5 5.1 1136 1 CED2_HUMAN Q96p48 homo sapien
40 86.5 5.1 2105 1 POLR_ASGVP P36309 apple stem
41 86.5 5.1 2805 1 MAPA_HUMAN P78559 homo sapien
42 86 5.1 436 1 TOLB_RHILQ Q98f84 rhizobium l
43 86 5.1 447 1 HSLD_PSEAE Q9huc5 pseudomonas
44 86 5.1 500 1 N1FD_RHICP P06769 rhizobium s
45 86 5.1 879 1 SYA_CLOPE Q8xjhe clostridium

```

#### ALIGNMENTS

```

RESULT 1
T2BB_BACSU STANDARD; PRT; 316 AA.
AC P33562; 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Type II restriction enzyme BsuBI (EC 3.1.21.4) (Endonuclease BsuBI)
DE (R_BsuBI).
GN HSDR OR HSRB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISB8;
RX MEDLINE=93126092; PubMed=1480472;
RA Xu G.-L., Kapfer W., Walter J., Trautner T.A.;
RT "BsuBI -- an isospecific restriction and modification system of PstI:
characterization of the BsuBI genes and enzymes.";
RL Nucleic Acids Res. 20:6517-6523(1992)
CC -!- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE CTGCAG AND
CLEAVES AFTER A-5.
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
specific double-stranded fragments with terminal 5'-phosphates.
CC -!- COFACTOR: Magnesium.
CC -!- SUBUNIT: Homodimer.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L01541; AAA18170.1; -.
DR PIR; S35516; S35516.
DR REBASE; 616; BsuBI.
KW Restriction system; Hydrolase; Nuclease; Endonuclease; Magnesium.
SQ SEQUENCE 316 AA; 36197 MW; B1C2E45851E7F3C9 CRC64;

```

```

Query Match 35.7%; Score 603.5; DB 1; Length 316;
Best Local Similarity 37.1%; Pred. No. 6.1e-40;
Matches 117; Conservative 65; Mismatches 132; Indels 1; Gaps 1;

QY 8 DGTVASIDTARALLKRFQDAQRYNVRSAVTLALAGLKPGRWVDSTTPRLGVQKIMDW 67
Db 3 EGMHNVKEAIKLGELPKQQNERSALCLSLMNITQDKTWEAESPLGIITPMMEF 62
QY 68 SEHWAKPYATGSRDFRKKTLRQWVDNGFAVLNADNLNIATNSQNLNCLSEALQALR 127
Db 63 CRINYGKEYAPNSRETFRRFTMHQFVDAGIALYNPDKPTRPVNSPKAVYQIEATLELTK 122
QY 128 AYTGFGEESLVVFLDEASKAVKARAEALQAAMISVDLPGGREFLLSPAGQNPLLKQWVE 187

```

```

Db      123  CYNTEWSELLARYLSNRQTLIVRYAKEROQNKIPVQIAEGKEIYITPGEHSELKAIIE 182
Qy      188  EFVPRFAPRSTVLYLGDTRGHSLFEREIFEVLGLTFDPHGRMPDLILHDEVGWLFLME 247
Db      183  EFAPRYVPGRLIYAGDTGERMGYFDEBLRQ-LGVWIDSHGKMPDVVYIPEKKWLLI 241
Qy      248  EAVSKGPFDEERHSLOELFVTPSAGLIFVNCENRESMQWLPELAWETEAWVAEDPD 307
Db      242  ESVTSHGPFVDRHREELAKLFGNGTAGIVYTAFPNRSLMARYLNNISWETEVWVADAPS 301
Qy      308  HLIHLNGSRFLGPYE 322
Db      302  HLIHFGVRFGLPYE 316

RESULT 2
ID T2PS_PROST STANDARD; PRT; 326 AA.
AC P06640;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Type II restriction enzyme PstI (EC 3.1.21.4) (Endonuclease PstI)
DE (R.PstI).
GN PSTIR.
OS Providencia stuartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxID=588;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=164;
RX MEDLINE=84239756; PubMed=6330092;
RA Walder R.Y., Walder J.A., Donelson J.E.;
RT "The organization and complete nucleotide sequence of the PstI
RT restriction-modification system."
RL J. Biol. Chem. 259:8015-8026(1984).
CC -!- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE CTCGAG AND
CC CLEAVAGES AFTER A-5.
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC specific double-stranded fragments with terminal 5'-phosphates.
CC -!- COFACTOR: Magnesium.
CC -!- SUBUNIT: Homodimer.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02081; AAA25673.1; -.
DR PIR; A00783; NDOFS.
DR REBASE; 1536; PstI.
KW Restriction system; Hydrolase; Nuclease; Endonuclease; Magnesium.
SQ SEQUENCE 326 AA; 37413 MW; 475841183B414504 CRC64;

Query Match 30.3%; Score 511; DB 1; Length 326;
Best Local Similarity 35.7%; Pred. No. 1.1e-32;
Matches 112; Conservative 53; Mismatches 141; Indels 8; Gaps 3;

Qy      17  ABALIKRFGDQQRNVRSVATLLALAGLKPGRWVDSVTPRLGVQKIMDWSGEHWAKPY 76
Db      9  AKEIILKALGPLPQVNDNRSGWLLALANIKPDSWKAKAPLLPTVSIMEFIRTEYKDY 68
Qy      77  ATGSRDFRKKTLROWVNGFVAVINADNLNIATNSQINCYCLSDALQALRAYGTGEPBE 136
Db      69  KNSRETIRQTLHQEQEQAIVDRNEDLPSRAINSKDNYSLSNQVIIDILHNPVNGNWE 128
Qy      137  SLIVFVLEASKAVKARAENALQAMISVDLPGGEEFLLSPAGQNPLLKQWVEEFVPRF-AP 195
Db      129  LIQQLFTHVPSLQELIYERALARDRIPKILDLDTGQISLSFGEHNLHADIVHFCRPFVD 188

```

```

Qy      196  RSTVLYLGDT- ----GKHSLEFEREIFEVLGLTFDPHGRMPDLILHDEVGWLFLME 249
Db      189  MGKILYIGDTASSRNEGGKMLVDSEYLLK-LGVPPMSHDKLPDVVYVDEKRWLFLEA 247
Qy      250  VKSGKPFDEERHSLOELFVTPSAGLIFVNCENRESMRQWLPELAWETEAWVAEDPD 309
Db      248  VTHGPISPKRWLEALSSCTVGKVVYVTAFTTRTFRKNAANIATWETVNIADNPDM 307
Qy      310  IHLNGSRFLGPYE 323
Db      308  VHFNGDRFLGPHDK 321

RESULT 3
VP4_ROTBA
ID VP4_ROTBA STANDARD; PRT; 772 AA.
AC P36307;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
GN S4.
OS Bovine rotavirus (serotype 10 / strain A44).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36437;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93286580; PubMed=8398807;
RA Taniguchi K., Urasawa T., Urasawa S.;
RT "Independent segregation of the VP4 and the VP7 genes in bovine
RT rotaviruses as confirmed by VP4 sequence analysis of G8 and G10
RT bovine rotavirus strains."
RL J. Gen. Virol. 74:1215-1221(1993).
CC -!- SUBCELLULAR LOCATION: Outer capsid.
CC -!- PTM: VP8 is one of two trypsin cleavage products of VP4; the other
CC product is VP5.
CC -!- SIMILARITY: Belongs to the rotaviruses VP4 protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13392; -. NOT ANNOTATED_CDS.
DR InterPro; IPR000416; Cap_VP4.
DR Pfam; PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 772 OUTER CAPSID PROTEIN VP4.
FT CHAIN 1 242 OUTER CAPSID PROTEIN VP8.
FT CHAIN 249 772 OUTER CAPSID PROTEIN VP5.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 772 AA; 86619 MW; D5FE1201624FE994 CRC64;

Query Match 6.2%; Score 104; DB 1; Length 772;
Best Local Similarity 22.9%; Pred. No. 2;
Matches 77; Conservative 47; Mismatches 128; Indels 84; Gaps 16;

Qy      1  VNSSDGIDGTVASIDTARALLKRFQDFDAQRYNVRSAVTLALA-----GLKPGDRWVDT 55
Db      473  VFSNDDYQTFPIANSVTVRQDLER-QLDEMRRREFNELSANIALSQLIDLALLP-----LDMF 527
Qy      56  TPRLLGVQKIMDWSGEHWAKPYATGSRDFRK-----KTLRQ 91

```

```
Db 528 SMFSGIOSTVE-----AAKTFATVSMKFKKSLAKSVNSLTDAITDAAGSISRSSTLRS 582
QY 92 -----WVNGFAVLNADNINATNSQLNECYCLSDALQALRAYGTEGP--EESLVVFL 142
Db 583 VNSVASVWTDLSIDVSTDNVVAAT-----ATAAKKFRVKEFTTEFNGVSFD 630
QY 143 DEASKAVKARAEALQAAMISVDLPFGGEEFLLSPAGQNPLLKKMVEEFPVPFAPR---STV 199
Db 631 DISAAVVKTKMKNLV---VD-----BELL-----PQITEASEKIPNRAVRLIDGK 676
QY 200 LYLGDTRKHSLSPEREIEFVEVLGTFPHGRMPDLILHDEVRGMLFLMEAVKS-KGPFDE 258
Db 677 VYEVTECKYFAYLTETFEVW---FDAE-RFAELVTDISOVISAIIDFKTKLNLDNYGI 732
QY 259 ERHRSLOELFVTPSAGLIFVNCENRSMROWLPEL 294
Db 733 TREOALNMLRSDPKVRSFIN--QNNPIKNRIEQL 766

RESULT 4
CARB STAEP STANDARD; PRT; 1057 AA.
ID CARB STAEP STANDARD; PRT; 1057 AA.
AC QCPJ4; 2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Ou D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AE016746; AAC04476.1; -.
DR HAMAP; MF_01210; -.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR000169; Shprot_acsite.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.
```

```
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF02142; MGS_1; -.
DR PRINTS; PRO0098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;

Query Match 6.0%; Score 101.5; DB 1; Length 1057;
Best Local Similarity 22.7%; Pred. No. 4.8;
Matches 61; Conservative 40; Mismatches 83; Indels 85; Gaps 14;

QY 82 EDRFKTKLROWVNDGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127
Db 214 KEIEYEVMDKNDNAIVVNCNMENIDPVGIHTGDSIVVAPSQTLSDVEYQMLRDVSLKVIR 273
QY 128 AYCTEG-----FEESLVVFLDEASKAVKARAEALQ-----NAMISVDLPGG 168
Db 274 ALGIEGCVNQLALDPSLNYIIIEVNPV--SRSSALASKATGYPIAKLAIAVGLTLD 332
QY 169 EEFLSP-----AGQNPLIKMVEEFVPRF-----APRSTVLYLGDT 206
Db 333 E--MLNPITGTSVAAFEPTLDYVISK-IPFPFDKPEKGERELGTOMKATGEVWAGRT- 388
QY 207 GKHSLEFERIFEVLGLTDPH-----GRMPDL-----ILHDEVRGMLFLMEAVKSKG 254
Db 389 -----YEESLLKAIRSLRYGVHHLGSLNGESYELDVIKERIGHODDERLFFICEAIR--- 440
QY 255 PFDEERHRSLOELFVTPSAGLIFVNCEN 283
Db 441 -----RGTSLLELHNMTKIDYFFLNKFN 464

RESULT 5
VP4_ROTBB
ID VP4 ROTBB STANDARD; PRT; 772 AA.
AC P35746;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
GN S4.
OS Bovine rotavirus (serotype 10 / strain B223).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10930;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033121; PubMed=1329320;
RA Hardy M.E., Gorziglia M., Woode G.N.;
RT "Amino acid sequence analysis of bovine rotavirus B223 reveals a
RT unique outer capsid protein VP4 and confirms a third bovine VP4
RT type.";
RL Virology 191:291-300(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93286580; PubMed=8389807;
RA Taniguchi K., Urasawa T., Urasawa S.;
```

RT "Independent segregation of the VP4 and the VP7 genes in bovine  
 RT rotaviruses as confirmed by VP4 sequence analysis of G8 and G10  
 RL bovine rotavirus strains.";  
 RL J. Gen. Virol. 74:1215-1221(1993).

CC -!- SUBCELLULAR LOCATION: Outer capsid.  
 CC -!- PTM: VP8 is one of two trypsin cleavage products of VP4; the other  
 CC product is VP5.  
 CC -!- SIMILARITY: Belongs to the rotaviruses VP4 protein family.  
 CC -----

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M92986; AAA02978.1; -

CC PIR; A44052; A44052.

CC PIR; JQ2025; JQ2025.

CC InterPro; IPR000416; Cap VP4.

CC InterPro; IPR008985; ConA\_like\_lec\_gl.

CC Pfam; PF00426; VP4; 1.

CC Coated protein; Glycoprotein.

CC CHAIN 1 772 OUTER CAPSID PROTEIN VP4.

CC CHAIN 1 242 OUTER CAPSID PROTEIN VP5.

CC CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CONFLICT 2 2 R -> A (IN REF. 2).

CC CONFLICT 193 193 N -> H (IN REF. 2).

CC CONFLICT 535 535 S -> T (IN REF. 2).

CC SEQUENCE 772 AA; 86540 MW; 61FA433F71079A8D CRC64;

Query Match 6.0%; Score 101; DB 1; Length 772;

Best Local Similarity 22.9%; Pred. No. 3.5;

Matches 77; Conservative 46; Mismatches 129; Indels 84; Gaps 16;

QY 1 VNSSDGDIGTVAIDTARALKRFGDQRYNVRSAVTLALA-----GLKPGDRWVDST 55

DB 473 VPSNDYQTPANSVTVQDLER-QLDEMRRFELSANIALSOLIDLALLP-----LDMF 527

QY 56 TPLRGVQKIMWSGEHNAKPYATGSRDPRK-----KTLRQ 91

DB 528 SMFSGIQSTVE-----AAKTFATSVNKKFRKSLAKSVNSLTDATDAAGSISRSTLRS 582

QY 92 -----WVDNGFAVLNADNLNIATNSQLNEYCLSDALQALRAYTGEF--EESLVVPL 142

DB 583 VNSAASVWTDISDIDVDSTDNVVAAT-----ATAAAKFRVKEFTTEFGVSD 630

QY 143 DEAKAVKARABALQAAMISVDLPGEFEPLSPAGONPLKKWVEEFVPRFAPR---STV 199

DB 631 DISAAVVKTKMKNLV-----VD-----EELI-----PQITEASEKFIENRAYRLIDGK 676

QY 200 LYLGDTRGKHSIFEREIFEIVLGLTFDPHGRMPDLILHDEVRGWLFLMEAVKS-KGPFDE 258

DB 677 VYEVTEGKYFAYLTETFEVW---FDAB-REAELVTDSPVISAIDFTIKNLNDNYGI 732

QY 259 EHRHSIQELFVTPSAGLIFVNCFNRESMROWLPEL 294

DB 733 TREQALNMLRSDPKVLRSPIN--QNNPIIKNRIEQL 766

RESULT 6

VP4\_ROTU

ID\_VP4\_ROTU

AC\_Q09113;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)  
 DE [Contains: Outer capsid proteins VP5 and VP8].  
 GN S4.

OS Human rotavirus (serotype G / strain 116E).

OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.

OX NCBI\_TaxID=42567;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93242780; PubMed=8386888;

RA Gentsch J.R., Das B.K., Jiang B., Bhan M.K., Glass R.I.;

RT "Similarity of the VP4 protein of human rotavirus strain 116E to that

RL of the bovine B223 strain.";

RL Virology 194:424-430(1993).

CC -!- SUBCELLULAR LOCATION: Outer capsid.

CC -!- PTM: VP8 is one of two trypsin cleavage products of VP4; the other

CC product is VP5.

CC -!- SIMILARITY: Belongs to the rotaviruses VP4 protein family.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L07934; AAA47357.1; -

CC PIR; A46110; A46110.

CC InterPro; IPR000416; Cap VP4.

CC InterPro; IPR008985; ConA\_like\_lec\_gl.

CC Pfam; PF00426; VP4; 1.

CC Coated protein; Glycoprotein.

CC CHAIN 1 772 OUTER CAPSID PROTEIN VP4.

CC CHAIN 1 242 OUTER CAPSID PROTEIN VP5 (BY SIMILARITY).

CC CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).

CC CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).

CC SEQUENCE 772 AA; 86882 MW; 31139CD20695F822 CRC64;

Query Match 5.9%; Score 100; DB 1; Length 772;

Best Local Similarity 22.6%; Pred. No. 4.2;

Matches 76; Conservative 48; Mismatches 128; Indels 84; Gaps 16;

QY 1 VNSSDGDIGTVAIDTARALKRFGDQRYNVRSAVTLALA-----GLKPGDRWVDST 55

DB 473 VPSNDYQTPANSVTVQDLER-QLDEMRRFELSANIALSOLIDLALLP-----LDMF 527

QY 56 TPLRGVQKIMWSGEHNAKPYATGSRDPRK-----KTLRQ 91

DB 528 SMFSGIRSTIE-----AAKNFATSVNKKFRKSNLAKSVNSLTDATDAAGSISRSTLRS 582

QY 92 -----WVDNGFAVLNADNLNIATNSQLNEYCLSDALQALRAYTGEF--EESLVVPL 142

DB 583 ANSASVWTDISDIDVDSTDNVVAAT-----ATAAAKFRVKEFTTEFGVSD 630

QY 143 DEAKAVKARABALQAAMISVDLPGEFEPLSPAGONPLKKWVEEFVPRFAPR---STV 199

DB 631 DISAAVVKTKMKNLV-----VD-----EEML-----PQITEASEKFIENRAYRLIDGK 676

QY 200 LYLGDTRGKHSIFEREIFEIVLGLTFDPHGRMPDLILHDEVRGWLFLMEAVKS-KGPFDE 258

DB 677 VYEVTEGKYFAYLTETFEVW---FDAB-REAELVTDSPVISAIDFTIKNLNDNYGI 732

QY 259 EHRHSIQELFVTPSAGLIFVNCFNRESMROWLPEL 294

DB 733 TREQALNMLRSDPKVLRSPIN--QNNPIIKNRIEQL 766











```

QY 169 BEFLSP-----AGQNPFLKKMVEEFVPRF-----APRSTVLYLGDT 206
Db 333 E--MNPITGTSYAAFEFLDYVISK-IPRPFDKFKGERELGTQMKATGEVMAIGRT- 388
QY 207 GKHSLEFEIPEEVLGLTFD-----PHGRMPDL-----ILHDEVRWGLFMLEAVKSG 254
Db 389 -----YEBSLLKAIRSLVGVHHLGLPNGESFDLDYIKERISHQDDERLFFIGEAIR--- 440
QY 255 PFDERHRSLOELFVTPSAGLIFVNCFFENRSMRWLPE 293
Db 441 -----RGTTLEIHNMTQIDYFFLHKFQNIIDIEHOLKE 474

RESULT 12
ID CARB_STAAM STANDARD; PRT; 1057 AA.
AC P58940;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR MW1086.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827 (2002).
CC -! CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -! COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -! PATHWAY: Arginine biosynthesis.
CC -! PATHWAY: Pyrimidine biosynthesis; first step.
CC -! SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -! SIMILARITY: Belongs to the carB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004825; BAB94951.1; -.
CC HAMAP; MF 01210; -.
CC InterPro; IPR006275; CarA_L_glu.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR004362; MGS like.
CC Pfam; PF02789; CPase_L_chain; 2.
CC Pfam; PF02786; CPase_L_D2; 2.
CC Pfam; PF02787; CPase_L_D3; 1.
CC Pfam; PF02142; MGS; 1.
CC PRINTS; PR00098; CPASE.
CC TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
CC PROSITE; PS00866; CPASE_1; 2.
CC PROSITE; PS00867; CPASE_2; 2.
CC ARGININE BIOSYNTHESIS; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.

```

```

FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9BC6F152 CRC64;

Query Match 5.7%; Score 95.5; DB 1; Length 1057;
Best Local Similarity 21.5%; Pred. No. 14;
Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;

QY 82 EDRKKTLRWVDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127
Db 214 KEIEYEVNRDKNDNAIVVCNMENIDPVGIHTGDSIVVAPSOQLSDVEYQMLRDVSLKVR 273
QY 128 AVYREG-----FEESLVVFLDEASKAVKARALQ-----AAMISVDLPGG 168
Db 274 ALGIEGGCNCVOLALDPHPSFDYIIIEVNPV--SRSSALASKATGYPIAKLAIAVGLTLD 332
QY 169 EEFLSP-----AGQNPFLKKMVEEFVPRF-----APRSTVLYLGDT 206
Db 333 E--MNPITGTSYAAFEFLDYVISK-IPRPFDKFKGERELGTQMKATGEVMAIGRT- 388
QY 207 GKHSLEFEIPEEVLGLTFD-----PHGRMPDL-----ILHDEVRWGLFMLEAVKSG 254
Db 389 -----YEBSLLKAIRSLVGVHHLGLPNGESFDLDYIKERISHQDDERLFFIGEAIR--- 440
QY 255 PFDERHRSLOELFVTPSAGLIFVNCFFENRSMRWLPE 293
Db 441 -----RGTTLEIHNMTQIDYFFLHKFQNIIDIEHOLKE 474

RESULT 13
ID PLEI_HUMAN STANDARD; PRT; 4684 AA.
AC Q15149; Q15149; Q16640;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HDI).
GN PLECI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Placenta;
RX MEDLINE=96210632; PubMed=8633055;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (8q24).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283 (1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.
RX MEDLINE=96312447; PubMed=8698233;
RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,
RA Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,
RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,
RA Uitto J.;
RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:
RT cDNA cloning and genomic organization.";
RL Genes Dev. 10:1724-1735 (1996).
RN [3]
RP VARIANT MD-EBS 1003-GLN->ALA-1005 DEL.
RX MEDLINE=97049959; PubMed=8894687;

```

RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
RA Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.,  
RA "Homozygous deletion mutations in the plectin gene (PLECI) in patients  
RT with epidermolysis bullosa simplex associated with late-onset  
RT muscular dystrophy.",  
RL Hum. Mol. Genet. 5:1539-1546(1996).  
RN [4]  
RP VARIANT MD-EBS LEU-429 INS.  
RX MEDLINE=21090821; PubMed=11151919;  
RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,  
RA Muss W., Hametner R., Klausberger A., Huber A., Pohl-Gubo G.,  
RA Wiche G., Uitto J., Hintner H.,  
RT "A compound heterozygous one amino-acid insertion/nonsense mutation in  
RT the plectin gene causes epidermolysis bullosa simplex with plectin  
RT deficiency.",  
RL Am. J. Pathol. 158:617-625(2001).  
RN [5]  
RP VARIANT EBS1 TRP-2110.  
RX MEDLINE=21841370; PubMed=11851880;  
RA Koss-Harnes D., Hoeyheim B., Anton-Lamprecht I., Gjesti A.,  
RA Joergensen R.S., Jahnsen F.L., Olaisen B., Wiche G.,  
RA Gedde-Dahl T. Jr.,  
RT "A site-specific plectin mutation causes dominant epidermolysis  
RT bullosa simplex Ogna: two identical de novo mutations.",  
RL J. Invest. Dermatol. 118:87-93(2002).  
CC -I- FUNCTION: Interlinks intermediate filaments with microtubules and  
CC microfilaments and anchors intermediate filaments to desmosomes or  
CC hemidesmosomes. Could also bind muscle proteins such as actin to  
CC membrane complexes in muscle. May be involved not only in the  
CC crosslinking and stabilization of cytoskeletal intermediate  
CC filament network, but also in the regulation of their dynamics.  
CC -I- SUBUNIT: Homodimer or homotetramer.  
CC -I- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC IsoId=Q15149-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q15149-2; Sequence=VSP\_005030;  
CC Name=3;  
CC IsoId=Q15149-3; Sequence=VSP\_005030, VSP\_005031;  
CC -I- TISSUE SPECIFICITY: Widely expressed with highest levels in  
CC muscle, heart, placenta and spinal cord.  
CC -I- DOMAIN: The N-terminus interacts with actin, the C-terminus with  
CC vimentin, desmin, GFAP, cyokeratins, lamin B; whereas both the N-  
CC and the C-terminus can bind integrin beta-4.  
CC -I- PTM: PHOSPHORYLATED BY CMC2; REGULATES DISSOCIATION FROM  
CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
CC -I- DISEASE: Defects in PLECI are the cause of epidermolysis bullosa  
CC simplex with muscular dystrophy (MD-EBS) [MIM:226670]; an  
CC autosomal recessive disorder characterized by epidermal blister  
CC formation at the level of the hemidesmosome and associated with  
CC late-onset muscular dystrophy.  
CC -I- DISEASE: Defects in PLECI are the cause of epidermolysis bullosa  
CC simplex 1 (EBS1) [MIM:131950]; also called epidermolysis bullosa  
CC simplex Ogna type. EBS1 is an autosomal dominant form of  
CC epidermolysis bullosa simplex differentiated from the more  
CC generalized form of Koebner [MIM:131900] and the localized form of  
CC Weber and Cockayne [MIM:131800] by the occurrence of skin  
CC bruising.  
CC -I- SIMILARITY: Contains 1 actin-binding domain.  
CC -I- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -I- SIMILARITY: Contains 33 plectrin repeats.  
CC -I- SIMILARITY: Contains 4 spectrin repeats.  
CC -I- SIMILARITY: Belongs to the plakin or cytolinker family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z54367; CAA91196.1; --  
DR EMBL; U53204; AAB05427.1; --  
DR EMBL; U63610; AAB05428.1; --  
DR EMBL; U63609; AAB05428.1; JOINED.  
DR EMBL; X97053; CAA65765.1; --  
DR PIR; C59404; A59404.  
DR HSSP; Q01082; 1BKR.  
DR Genew; HGNC:9069; PLECI.  
DR GK; Q15149; --  
DR MIM; 601282; --  
DR MIM; 226670; --  
DR MIM; 131950; --  
DR GO; GO:0008307; F:structural constituent of muscle; TAS.  
DR InterPro; IPR001589; Actbind actnin.  
DR InterPro; IPR001715; Galponin-like.  
DR InterPro; IPR001101; Plectrin repeat.  
DR InterPro; IPR005326; S10\_plectin\_N.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00307; CH; 2.  
DR Pfam; PF00681; Plectin; 19.  
DR Pfam; PF03501; S10\_plectin; 1.  
DR ProDom; PD006662; S10\_plectin\_N; 1.  
DR SMART; SM00033; CH; 2.  
DR SMART; SM00250; PLEC; 34.  
DR PROSITE; PS00019; ACTININ\_1; FALSE NEG.  
DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
DR PROSITE; PS00021; CH; 2.  
KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;  
KW Disease mutation.  
FT DOMAIN 1 1470 GLOBULAR 1.  
FT DOMAIN 1471 2755 CENTRAL FIBROUS ROD DOMAIN.  
FT DOMAIN 2756 4684 GLOBULAR 2.  
FT DOMAIN 175 400 ACTIN-BINDING.  
FT DOMAIN 179 282 CH 1.  
FT DOMAIN 295 397 CH 2.  
FT REPEAT 645 710 SPECTRIN 1.  
FT REPEAT 740 824 SPECTRIN 2.  
FT REPEAT 837 930 SPECTRIN 3.  
FT REPEAT 1315 1415 SPECTRIN 4.  
FT DOMAIN 1469 2756 COILED COIL (POTENTIAL).  
FT REPEAT 2826 2863 PLECTIN 1.  
FT REPEAT 2864 2901 PLECTIN 2.  
FT REPEAT 2902 2939 PLECTIN 3.  
FT REPEAT 2940 2977 PLECTIN 4.  
FT REPEAT 2981 3015 PLECTIN 5.  
FT REPEAT 3116 3153 PLECTIN 6.  
FT REPEAT 3154 3191 PLECTIN 7.  
FT REPEAT 3192 3229 PLECTIN 8.  
FT REPEAT 3230 3267 PLECTIN 9.  
FT REPEAT 3268 3305 PLECTIN 10.  
FT REPEAT 3306 3343 PLECTIN 11.  
FT REPEAT 3485 3522 PLECTIN 12.  
FT REPEAT 3523 3560 PLECTIN 13.  
FT REPEAT 3561 3598 PLECTIN 14.  
FT REPEAT 3599 3636 PLECTIN 15.  
FT REPEAT 3640 3674 PLECTIN 16.  
FT REPEAT 3820 3857 PLECTIN 17.  
FT REPEAT 3858 3895 PLECTIN 18.  
FT REPEAT 3896 3933 PLECTIN 19.  
FT REPEAT 3934 3971 PLECTIN 20.  
FT REPEAT 3975 4008 PLECTIN 21.  
FT REPEAT 4063 4100 PLECTIN 22.  
FT REPEAT 4101 4138 PLECTIN 23.  
FT REPEAT 4139 4176 PLECTIN 24.  
FT REPEAT 4177 4214 PLECTIN 25.  
FT REPEAT 4218 4252 PLECTIN 26.  
FT REPEAT 4265 4305 PLECTIN 27.  
FT REPEAT 4319 4356 PLECTIN 28.  
FT REPEAT 4408 4445 PLECTIN 29.  
FT REPEAT 4446 4483 PLECTIN 30.  
FT REPEAT 4484 4521 PLECTIN 31.  
FT REPEAT 4522 4559 PLECTIN 32.

```

FT REPEAT 4560 4597 PLECTIN 33.
FT DOMAIN 4250 4300 BINDING TO INTERMEDIATE FILAMENTS
FT 4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT 4625 4640 (BY SIMILARITY).
FT MOD_RES 4539 PHOSPHORYLATION (BY CDC2)
FT 1 174 (BY SIMILARITY).
FT VARSPLIC 1 174 MVAGMLPRDOLRAIYELVLFREGVVMVAKKRRPRSLHPHP
FT VTNLQVMAWASLRARGLVRETFAWCHFFWYLTNVEGTAHL
FT RQYLHLPEIYVAAQLORVRPVAMPARRPHVQAVQGPL
FT GSPPKRGPELPTEQRLYRKELEEVSPETVPATTQRTLA
FT RPPGPAPAT -> MSGEDAEVAVSDVSSGSPSPGD
FT TILPWLNGTKQSRSGGAGNGSVLDPAERAVIRIA (in
FT isoform 2 and isoform 3).
FT /FTId=VSP 005030.
FT Missing (in isoform 3).
FT /FTId=VSP 005031.
FT L -> LL (in MD-EBS).
FT /FTId=VAR 011336.
FT Missing (in MD-EBS).
FT /FTId=VAR 011337.
FT R -> W (in EBS1).
FT /FTId=VAR_015817.

```

```

Query Match 5.6%; Score 95; DB 1; Length 4684;
Best Local Similarity 25.9%; Pred. No. 1.1e+02;
Matches 62; Conservative 39; Mismatches 70; Indels 68; Gaps 14;

QY 84 FRKTKLQ--WVDNGFAVINADNLNI--ATNSQLNEYCLSDALQALRAYGTGFE-- 136
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2310 FAEQTLRQAQVEQLTTLT---RLQLEETDHQK---LLDELOLKLKAEATEAARQSQV 2363

QY 137 -----SLVFLDEASKAVKARAEALQAAISVDLPGEFFLLSPAGONPLKMKVVEEFPV 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2364 BEELFSVRQMEELSK-LKARIEANRALILDKONTORFLOEBAK---MKQVAAE--- 2416

QY 192 RPAPSTVLYLGDTRGKHSLEFEIFEVLGTFDPHGMRPDLILHDEVRGWLFLMEAVK 251
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2417 ---AARLSV-----AAQEAARLQLAEDLA---QORALAEKMLKEMQA---VQEA 2461

QY 252 SKGPPD-----BEHRSLQELFVTPSAGLIFVNCFFENRESRWMLPELAWTEAW 301
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2462 LKAAEALLQOKELAQEQARRLQ-----EDKEQMAQ---QLAAETQGF 2501

```

```

RESULT 14
TOLB_BRAJA
ID TOLB_BRAJA STANDARD; PRT; 454 AA.
AC Q89ED5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE TOLB protein precursor.
GN TOLB OR BHL7152.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RT DNA Res. 9:189-197(2002).
CC -!- FUNCTION: Involved in the tonB-independent uptake of proteins (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to the tolB family.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP005961; BAC52417.1; ALT_INIT.
DR HANAP; MF_00671; ; 1.
DR InterPro; IPR006311; Tat.
DR InterPro; IPR007195; TolB_N.
DR Pfam; PF04052; TolB_N; 1.
DR TIGRFAMs; TIGR01409; Tat signal seq; 1.
KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 30 Potential.
FT CHAIN 31 454 TolB protein.
FT SEQUENCE 454 AA; 49857 MW; 522EB7A845EEDC78 CRC64;
SQ
Query Match 5.6%; Score 94.5; DB 1; Length 454;
Best Local Similarity 26.4%; Pred. No. 5.7;
Matches 47; Conservative 23; Mismatches 53; Indels 55; Gaps 9;

QY 116 YCLSDALQALRAYGTGFEESLVVFLDEA---SKAVKARAEALQAAISVD-----LP 166
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 158 HIISDQIYE--RMTGKGYFDSRVVVDGTGPKERRVK-----RLAMMDQDQGANVRLT 209

QY 167 GGEFEELSPAGONPLKMKVVEEFPVPRAPS---TVLYLGDTRGKHSLE---EREIFE 218
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 210 RGSDDLVT-----PRFSPQEIITYMEFGQDPPKVLNFYNTGTGREIVG 253

QY 219 EVLGLT----FDPHGMRPDLILHDEVRGWLFLMEAVKSGPFDERRHRSLSQLFLVTPS 272
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 254 NFPGMTFAPRSPDQGVIMSLQGGNSNLFVMDL-----RSRSTRLTDTPA 301

RESULT 15
Y817_ARCFU
ID Y817_ARCFU STANDARD; PRT; 619 AA.
AC Q29441;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF0817 precursor.
GN AF0817.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: TO M.JANNASCHII MJ1554.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR -----  
DR EMEL; AF001048; AAB90425.1; --  
DR FIR; A69352; A69352.  
DR TIGR; AF0817; --  
KW Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 619 HYPOTHETICAL PROTEIN AF0817.  
SQ SEQUENCE 619 AA; 70225 MW; 8F13976711E7B24E CRC64;

Query Match 5.6%; Score 94.5; DB 1; Length 619;  
Best Local Similarity 21.5%; Pred. No. 8.5;  
Matches 82; Conservative 50; Mismatches 109; Indels 141; Gaps 19;

QY	13	SIDTARALLKFFGDAQRYNVRSAVTLLALGLK-----PGDRWVD-----	53
Db	249	TVDTTYTWK---LDAKTGEVENSFVGSGMSVVTYNSKNAIYVTYNSYADPAKLTQF	305
QY	54	-STTP-----RLGVOKIMDWSGEHWAKPY-----ATGSR-----	81
Db	306	ISENPDLVPDWIREKIEKLMYDISSRAKQVEIMYLLEQLRASWSEDERLKPFENEYNRW	365
QY	82	EDFRKKTILROWDNGFA----VLNADNLNIATNSQLNEYCLSDRALQALRAYGTEGFEES	137
Db	366	ENFTKHARETEKTHIAKFSIQLEAEGMNSVPGRLNRFSL-DEYNGYLRVAITVDWDEN	424
QY	138	LVVPLDEASKAV-KARAEALQAMISVDLPGEFEFLSPAG-----ONPLLKQKVE	187
Db	425	DLVYLDKLEWVGKIQGFGLDERIYAVRFDGVDGFIYTRQTPDFVFLDLSNPENPKIVG	484
QY	188	EF-VPRFA-----PRSTVLVYLGDTGR--XHSLFE-----REIFEVL--	221
Db	485	ELKIPGSSYLHRIDENTVLGVGREGNVKLSLFDISDLTSPKEKNRYILQETWSEVLSN	544
QY	222	-----GLTFDPHGRMPDLILHDEVGWLF-----LMEAVKSK-----	253
Db	545	HHAFLLDSQHGIFFLPAGQ-----NGYIFSYKDGKLIKAVKGNVRAIYIDDL	594
QY	254	---GP-----FDEERHRSLOEL	267
Db	595	YIIGPEEISYIDENSWEKVGEL	616

Search completed: October 1, 2004, 16:13:21  
Job time : 10.5859 secs

Blank



```
QY 122 ALQALRAYGTGPFESLVVFLDEASKAVKARAEALQAAMISVDLPQGEFFLLSPAGQNPL 181
Db 343 FVTLLKSYSGQWEETRNNYVISVKNLLQDRNNI--PMIPVSLPNGQAIQLSSGGQNIL 400
QY 182 LKKWVEEFVPRFAPRSTVLYLGDTRGKHSIFERIEFEEVLGLTFDPHGRMPDLIIHDEVR 241
Db 401 IKEILENFCPRFTPEGVLVFGDAGNKFIIVNETQKFEI-GIELDPHGKMPDIVVYERQ 459
QY 242 GWLFIMEAVKSGPPDEERHRSLOELFVTPSAGLIFVNCFNRRSMRWLPELAWETAW 301
Db 460 EWLVLIAVTTGHPVNLKRRNELKLFQSSRQGLVFVTAFFPSRKEMTRYLAELISMETEVM 519
QY 302 VAEDPDHLIHLNGSRFLGPE 322
Db 520 VAAQPDHMIHNGERFLGPE 540

RESULT 2
Q8VLS0 PRELIMINARY; PRT; 826 AA.
ID Q8VLS0
AC Q8VLS0;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Methyltransferase-endonuclease.
GN YEN1.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSOM=IS1222;
RA Rakin A.V.;
RT "Characterization of the restriction-modification system of Y.
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RE EMBL; AJ414030; CAC95150.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR000051; SAM bind.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Endonuclease.
SQ SEQUENCE 826 AA; 93348 MW; 148432FD8E798843 CRC64;

Query Match 35.8%; Score 604.5; DB 2; Length 826;
Best Local Similarity 39.3%; Pred. No. 7.1e-41;
Matches 121; Conservative 58; Mismatches 128; Indels 1; Gaps 1;

QY 14 ITARALKRFGFDAQRYNRSVAVILLALAGLKPGDRWVDSVTPRLGVQKIMDSGEHWA 73
Db 519 LQQALEIITLLGMPSQQNRSALTFLALVNLRPESGQWLEKPLVGVTPIMDCRDIYG 578
QY 74 KYATGSRDPRKKTLRQWVDNGFVNLNADNLNATNSQLNEYCSDEALQALRAYGTG 133
Db 579 KEYANTRETFRQTLHQFIDGLVLYNPKFNRAVNSPKACYQIAPELFDVLYNTYGTPL 638
QY 134 FRESLVVFLDEASKAVKARAEALQAAMISVDLPQGEFFLLSPAGQNPLKKWVEEFVPRF 193
Db 639 MNKALGEWLWQRETLVEQYAKREMHMIPLITIDNGTEIHLSPGDHSQLIHDIVTEFGPRF 698
QY 194 APRSTVLYLGDTRGKHSIFERIEFEEVLGLTFDPHGRMPDLIIHDEVRGWLFIWEAVKSK 253
Db 699 AFGSQVILYLGDTGAKEDDFRKLALAD-LGVTVNRKGLPDVVLVWPDQWNLILIVTSIH 757
```

```
QY 254 GPFDEERHRSLOELFVTPSAGLIFVNCFNRRSMRWLPELAWETAWVABDPDHLIHLN 313
Db 758 GPFVKGKHSSELNLFKDAARPGLVVVSAPPDKTKWSPFSEISWETEVWIAEAPTHMIHLN 817
QY 314 GSRFLGYPY 321
Db 818 GDREFLGPH 825

RESULT 3
Q9RBJ6 PRELIMINARY; PRT; 364 AA.
ID Q9RBJ6
AC Q9RBJ6;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Restriction endonuclease homolog R.Xphi.
GN XPHIR.
OS Xanthomonas campestris (pv. phaseoli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=29445;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xcp 73;
RA Lai J.-Y., Yang M.-T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF042157; AAF22367.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
KW Endonuclease.
SQ SEQUENCE 364 AA; 39725 MW; 8D150D6F40069A3D CRC64;

Query Match 14.2%; Score 240.5; DB 2; Length 364;
Best Local Similarity 27.2%; Pred. No. 2.1e-11;
Matches 94; Conservative 56; Mismatches 118; Indels 77; Gaps 18;

QY 7 IDGT-----VASIDTARALLKRFQDAQRYNRSVAVILLALAGLKPGDRWVDSVTPR 58
Db 48 VEGTGRWLAPKXHMVMTVEALLSD---DA-----VREA---YGLGAMKPGFR----- 89
QY 59 LGVQKIMDSGEHWAQPYATGSRDFRKTTLROWDNGFAVLN--ADNLNATNSQLNEY 116
Db 90 -----VAGQRW---YEENSREPLRDLRQ-----GFTTNVAERTGLFTTGLPRY 134
QY 117 CL-----SDEALQALRAYGTGFEESLVVFLDEASKAVKARAEALQAAMISV 163
Db 135 ALKTDFAALFDPALAGDDLIATAWQEAHLASAM-----ARIALVRGAAPTDEGVW 189
QY 164 DLPQGEFFLLSPAGQNPLKKWVEEFVPRFAPRSTVLYLGDTRGKHSIFERIEFEEV-IG 222
Db 190 TFPNGETRRMAPGPSVISKAVIEEFAARFLTQPAVLWVSESAKVVSRRDDELAASKLK 249
QY 223 LTFDPHGRMPDLIIHD-----EVRGWLFL-MEAVKSGPFDDEERHRSLOELFV----TSA 273
Db 250 ITADRN--LPDILVLDGGGVTFGLLVFIEVATDGPITMQRREAFWQIAADAGFTPEA 307
QY 274 GLIFVNCFNRR--SMRWLPELAWETAWVABDPDHLIHL-NGS 315
Db 308 -VAFVATYILDRSHTAFKKTIAELANRSFAWFASEPEHIIALNGA 351

RESULT 4
Q52863 PRELIMINARY; PRT; 353 AA.
ID Q52863
AC Q52863;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Endonuclease.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=384;
```

```

[1]
SEQUENCE FROM N.A.
RX MEDLINE=9805154; PubMed=3939436;
RX Rochepeau P., Selingue L.B., Hynes M.F.;
RA "Transposon-like structure of a new plasmid encoded restriction-
RT modification system in Rhizobium leguminosarum VF39SM.";
RL Mol. Gen. Genet. 256:387-396(1997).
DR EMBL; X99520; CAA67875.1; -.
SQ SEQUENCE 353 AA; 38174 MW; C74CF7DC8D011A7B CRC64;

Query Match      12.3%; Score 207; DB 2; Length 353;
Best Local Similarity 24.0%; Pred. No. 1.2e-08;
Matches 82; Conservative 51; Mismatches 145; Indels 64; Gaps 12;

Qy 1 VNSDSDGDTGVASIDTARALLKRGFGAQRNVRSVTLALAGLKPGDRWVDSSTPRLG 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 46 INAVDGGSIQLAPKHVYRMTQEQSLITDAHRLAYATAIMKAGHVDSGRW----- 96

Qy 61 VQKIMDWSGEHWAKPYATGSRGPRKXTLRQWDNGFAVLNADNLNIATNSQLNEY---- 116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 -----YLDNTRTSIRDDTSREALVATGAVI--EDTTAATTSKGKRYALQA 139

Qy 117 -----CLSDREAL-QALR-----AYTEGFEEISLVFLDPAKAVKARABALQAAMISV 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 AFAALMVPALQDEALDQAINENWRGXYITAGALARVAI-----VRAGA-ATGCTHVQV 190

Qy 164 DLPGGEFFLLSPAGONPLLLKKWVEEFPVPFAPRSTVLVGLDTRKGKHSILFEEI FEEVLGL 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 TFPNGETTRILKAGPSSVITKDVIEVFSPRFLGDPFAVLVFSES-GNKVVARDEKLTATIGL 249

Qy 224 TFDPHGRMPDLIILHD--EVRGWLFLMEAVKSGPFDDEHRHSLOELFFVTPSAG-----LI 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 SIQSDKDLDPDIIILVDLKPAPHELLVFEVWATDGPVGVRRAALEBK--LAEDAGFDLQHVA 307

Qy 277 FVNCFNRR--ESMRQWLPDLAWETEAVVADPDHLIHLNGSR 316

Db 308 FVTAYLDRSOSTFKRTAETLAWGTFAWFGAPPEHIVELSEGR 349

```

```

RESULT 5
Q8NLY2
ID Q8NLY2 PRELIMINARY; PRT; 1111 AA.
AC Q8NLY2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Cgl2804.
DE CGL2804.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
[1]
SEQUENCE FROM N.A.
RN STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032." ;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005283; BAC00198.1; -.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 1111 AA; 123962 MW; E354FB360EDDF9FA CRC64;

```

```
Query Match          7.0%; Score 118; DB 16; Length 1111;
Best Local Similarity 25.5%; Pred. No. 1.4;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;
```

Qy	131	TGFEESUUVLD-----EASKAVKARAEALQAAMISVDLPGBEEF 171
Dd	686	DVESADAIAELDRLELNNTPEATELSARHEAAKQTILRVSDLIIVAAQSEETVASMNL 745
Qy	172	LSPAGQNPLIKKWVEEPVPAPRSTVLYLGDTRGKHSLPEREFEEVLGLTFD 226
Dd	746	KRAETELKRUESLPVAEVSEIAREVEKLFLANTERVHA--ANVDEQTIARED 797
	RESULT 6	
ID	Q9JUL55	
AC	Q9JUL55;	PRELIMINARY; PRT: 331 AA.
DT	01-OCT-2000	(TrEMBLrel. 15, Created)
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)
DE		Membrane interacting protein of RGS16.
GN	MIR16.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
EN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=20226047; PubMed=10760272;	
RZ	Zhang B., Chen D., Farquhar M.G.;	
RT	"MIR16, a putative membrane glycerophosphodiester phosphodiesterase,	
RT	interacts with RGS16.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 97:3999-4004(2000) .	
DR	EMBL: AF212851; AAF55233.1;	
DR	GO: GO:0008899; F:glycerophosphodiester phosphodiesterase act. . ; IEA.	
DR	GO: GO:0006071; P:glycerol metabolism; IEA.	
DR	InterPro: IPR004129; GDDP.	
DR	Pfam: PF03009; GDDP; 1.	
SQ	SEQUENCE 331 AA; 37633 MW; D946067B5766E35F CRC64;	

Query Match	6.1%;	Score 103;	DB 11;	Length 331;	
Best Local Similarity	23.0%;	Pred. No. 3.9;			
Matches	78;	Conservative 36;	Mismatches 95;	Indels 130;	Gaps 19
Qy	20	LLKRFGDAQRNNVRS	AVTLLALAGLKPGDRWDVSTT	PRLGQVKIMDSGEH	----- 71
Db	41	LLRFFSFE	-----	PVPSRRALQVLKPRDR	-----VSAIAHRGGSHDAPENTUA 83
Qy	72	-----	WAKPYATGSRREDPRFKKTL	ROWDNGFAVLNADN	-----L 105
Db	84	AIRQAAXKGATGVELDI	-----	EFTSDGVPVLAHMDNTVDR	TTDGSGRGLCDLTFEQVRKL 137
Qy	106	NIATNSOL-NEYCUSD	EALQALRAYGTGFPESLVVFL	DEASKAVKAPAEALQAA	MSVD 164
Db	138	NPAANHRLNEF	-----	PDERIPTLREAVTECLCHN	ITIFFD-----VKGHADMASAAKNTY 190
Qy	165	LPGGEFEFLSPAGQNPLL	-----	KKMVEEFVPR	-----FAPRST-----VLVLGDR 206
Db	191	M-----EF	-----	POLYNNSMVCSFLPEVIY	KMRQTDQKVITALTHTRPWSLSHTGDGK 238
Qy	207	GKHSLFEREIPEEVLG	TFD--PHG-----	RMPDILIHDEVRGVMFL	MEAVKS 252
Db	239	PRYSVFVKQSFVFLD	ILLDWSMENVLMYLGCIS	AFMLQKDFVSPDYLLKKW	-----SA 291
Qy	253	KG-----	PPDEERH--	RSLOELFVTPSAGLIFVNC	280
Db	292	KGIQVSWTNTPTDEK	NYYESHLGSSYITDS	---MLENC	327
RESULT 7					
Q819U2					
ID	Q819U2		PRELIMINARY;	PRT;	444 AA.
AC	Q819U2;				
DT	01-JUN-2003	(TrEMBLrel. 24, Created)			
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)			



16S rRNA m(5)C 967 methyltransferase (EC 2.1.1.-).  
 DE BC3863.  
 GN Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=226900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapachal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,  
 RA Greckin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyrpides N.;  
 RT "genome sequence of Bacillus cereus and comparative analysis with  
 RT Bacillus anthracis.";  
 RL Nature 423:87-91(2003).  
 DR EMBL; AE017010; AAP10785.1; --.  
 DR GO; GO:0003723; F-RNA binding; IEA.  
 DR GO; GO:0008757; F-S-adenosylmethionine-dependent methyltransf. .; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR006027; NueB.  
 DR InterPro; IPR006174; NueB dom.  
 DR InterPro; IPR000051; SAM bind.  
 DR InterPro; IPR004573; Sun.  
 DR InterPro; IPR001678; Sun\_Nop1/Nop2.  
 DR Pfam; PF01189; Noll\_Nop2; Sun.  
 DR Pfam; PF01029; NueB; 1.  
 DR ProDom; PD005242; NueB dom; 1.  
 DR TIGRFAMs; TIGR00563; rsmB; 1.  
 DR PROSITE; PS01153; NOLL NOP2 SUN; 1.  
 DR Methyltransferase; Transferase; Complete proteome.  
 KW Methyltransferase; Transferase; Complete proteome.  
 SQ SEQUENCE 444 AA; 49940 MW; 7AF445EC3689D08C CRC64;  
 Query Match 6.1%; Score 102.5; DB 16; Length 444;  
 Best Local Similarity 21.5%; Pred. No. 6.7;  
 Matches 72; Conservative 50; Mismatches 140; Indels 73; Gaps 17;  
 QY 12 ASIDTARALLKRGFGDAQYNNVRSAYTLALAGLPGDRWVDSTPRLGVQKIMD-WSGE 70  
 DB 95 AALHEAVEAKRRGHGIGAGWNGVLRSTQRGVSVDE-KDPPERLAITTSHPWLQ 153  
 QY 71 HWAKPYATGSRDFRKKTLROWVDNGFAVLNADNLNI-----ATNSQLNEYCLSD 120  
 DB 154 EWTSEYGLETAEKWCVNMLPPVPT--ARVNVKVTVEAEIILLASEGIEAKLGD--LSD 209  
 QY 121 EALQALR--AYGTGEGEESLVVFLDEASKAVKARAEALQ-----AAMISVDLPGEFLL 173  
 DB 210 DAIQIERGNVAHTFAFKGFLSTQDESSMLV---ARALEPNTGDAVLDSCAAPGKTHI 266  
 QY 174 SP--AGQNPLK-----KWEEFVPFAPRSTVLYIGDTRGKHSLFEREIEEVL-- 221  
 DB 267 AERLKGKGVMSLDLHAHKVRLIKQAERLGLNENVETKALDARKVQEHAFNETDKILVD 326  
 QY 222 ----GLTFDPHGRMPDLILHDEVRGWLFLMEAVKSGPDEHRHSLQELFVTPSA----- 273  
 DB 327 APCSG--FGVIRKPKDKLG-----KDKG--DSERLSTIQIATLEKTAPLKL 369  
 QY 274 ---GLIFVNC-----FENRESMOWL---PELAWET 298  
 DB 370 QGRELIVYSTCTIEKNEQVIEKLEHFEFENDT 404  
 RESULT 8  
 ID Q8EQH8 PRELIMINARY; PRT; 671 AA.  
 AC Q8EQH8:  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Copper-transporting Arpase (EC 3.6.1.-).  
 GN O81721.  
 OS Oceanobacillus ihevensis.  
 OX NCBI\_TaxID=70813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22608415; PubMed=12721630;  
 RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,  
 RA Kapachal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,  
 RA Greckin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyrpides N.;  
 RT "genome sequence of Bacillus cereus and comparative analysis with  
 RT Bacillus anthracis.";  
 RL Nature 423:87-91(2003).  
 DR EMBL; AE017010; AAP10785.1; --.  
 DR GO; GO:0003723; F-RNA binding; IEA.  
 DR GO; GO:0008757; F-S-adenosylmethionine-dependent methyltransf. .; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR006027; NueB.  
 DR InterPro; IPR006174; NueB dom.  
 DR InterPro; IPR000051; SAM bind.  
 DR InterPro; IPR004573; Sun.  
 DR InterPro; IPR001678; Sun\_Nop1/Nop2.  
 DR Pfam; PF01189; Noll\_Nop2; Sun.  
 DR Pfam; PF01029; NueB; 1.  
 DR ProDom; PD005242; NueB dom; 1.  
 DR TIGRFAMs; TIGR00563; rsmB; 1.  
 DR PROSITE; PS01153; NOLL NOP2 SUN; 1.  
 DR Methyltransferase; Transferase; Complete proteome.  
 KW Methyltransferase; Transferase; Complete proteome.  
 SQ SEQUENCE 444 AA; 49940 MW; 7AF445EC3689D08C CRC64;  
 Query Match 6.1%; Score 102.5; DB 16; Length 444;  
 Best Local Similarity 21.5%; Pred. No. 6.7;  
 Matches 72; Conservative 50; Mismatches 140; Indels 73; Gaps 17;  
 QY 12 ASIDTARALLKRGFGDAQYNNVRSAYTLALAGLPGDRWVDSTPRLGVQKIMD-WSGE 70  
 DB 95 AALHEAVEAKRRGHGIGAGWNGVLRSTQRGVSVDE-KDPPERLAITTSHPWLQ 153  
 QY 71 HWAKPYATGSRDFRKKTLROWVDNGFAVLNADNLNI-----ATNSQLNEYCLSD 120  
 DB 154 EWTSEYGLETAEKWCVNMLPPVPT--ARVNVKVTVEAEIILLASEGIEAKLGD--LSD 209  
 QY 121 EALQALR--AYGTGEGEESLVVFLDEASKAVKARAEALQ-----AAMISVDLPGEFLL 173  
 DB 210 DAIQIERGNVAHTFAFKGFLSTQDESSMLV---ARALEPNTGDAVLDSCAAPGKTHI 266  
 QY 174 SP--AGQNPLK-----KWEEFVPFAPRSTVLYIGDTRGKHSLFEREIEEVL-- 221  
 DB 267 AERLKGKGVMSLDLHAHKVRLIKQAERLGLNENVETKALDARKVQEHAFNETDKILVD 326  
 QY 222 ----GLTFDPHGRMPDLILHDEVRGWLFLMEAVKSGPDEHRHSLQELFVTPSA----- 273  
 DB 327 APCSG--FGVIRKPKDKLG-----KDKG--DSERLSTIQIATLEKTAPLKL 369  
 QY 274 ---GLIFVNC-----FENRESMOWL---PELAWET 298  
 DB 370 QGRELIVYSTCTIEKNEQVIEKLEHFEFENDT 404  
 RESULT 9  
 ID Q7X5B0 PRELIMINARY; PRT; 496 AA.  
 AC Q7X5B0:  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Dinitrogenase alpha subunit.  
 GN N1FD.  
 OS Fischerella sp. UTEX 1903.  
 OC Bacteria; Cyanobacteria; Stigonematales; Fischerella.  
 OX NCBI\_TaxID=70813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama I.;  
 RT "genome sequence of Oceanobacillus ihevensis isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.";  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL; AF004598; BAC13677.1; --.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0008524; F:ATP binding; IEA.  
 DR GO; GO:0004008; F:copper-exporting ATPase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0006825; P:copper ion transport; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR GO; GO:0030001; P:metal ion transport; IEA.  
 DR InterPro; IPR006403; ATPase-IBI Cu.  
 DR InterPro; IPR006416; ATPase-IBI hvy.  
 DR InterPro; IPR001757; ATPase\_EI-E2.  
 DR InterPro; IPR008250; EI-E2 ATPase.  
 DR InterPro; IPR005834; Hydrolase.  
 DR Pfam; PF00122; EI-E2 ATPase; 1.  
 DR Pfam; PF00702; Hydrolase; 1.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PRINTS; PR00943; CUATPASE.  
 DR TIGRFAMs; TIGR01511; ATPase-IBI Cu; 1.  
 DR TIGRFAMs; TIGR01525; ATPase-IBI hvy; 1.  
 DR PROSITE; PS00154; ATPASE EI E2; 1.  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 671 AA; 73122 MW; F2C3673246F02881 CRC64;  
 Query Match 6.1%; Score 102.5; DB 16; Length 671;  
 Best Local Similarity 18.0%; Pred. No. 12;  
 Matches 48; Conservative 49; Mismatches 82; Indels 87; Gaps 9;  
 QY 65 MOWSGHWAQ-----PVATGSRDFRKKTLROWVDNGFAVLNADNLNIATNS 111  
 DB 62 LFSFGHWEVFILATVIFYGWPFITGLFDEIKOKSPGMTLIGFAIVAVIYSAAT-- 119  
 QY 112 QLNEYCLSEALQALRAYTGEFE-----ESLWVF-----LDEASKAVKARAE 155  
 DB 120 -----VFGIDGNLFWELATLIVIMLLGHWIEKMSINKASESEALADL 163  
 QY 156 LQAAMISVDLPG-----GEELLSPAGNPLKKWVEFVPRFAPRSTVLYLG 203  
 DB 164 MQEAVKIDTEGNTETISISHLQTDGHLIKPEK-----VPDAVIK 208



```
GN SE0879,
OS Staphylococcus epidermidis
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016746; AA004476.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004086; F:carbamoyl-phosphate synthase activity; IEA.
DR GO; GO:0004137; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0006526; P:arginine biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR004362; MGS-like.
DR InterPro; IPR000169; SphroT_acsite.
DR Pfam; PF00289; CPSase_L.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
DR Complete proteome.
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DBICAE59 CRC64;

Query Match 6.0%; Score 101.5; DB 16; Length 1057;
Best Local Similarity 22.7%; Pred. No. 30;
Matches 61; Conservative 40; Mismatches 83; Indels 85; Gaps 14;

QY 82 EDFRKKTLRQWVNGFAVLNADNL-----IATNSQLN--EY-CLSDALQALR 127
Db 214 KEIYEVRKDKNDNAIVVCCNENIDPVGIHGTDSIVVAPSTQLSDVEIQMLRDSLKVIR 273
QY 128 AYGTEG-----FEESLVVFLDEASKAVKARAEALQ-----AAAMISVDLPFG 168
Db 274 ALGIEGGCNVQLADPHSLNYVYIIEVNPV-SRSSALASKATGYPIAKLAIAVGLTLD 332
QY 169 EEFLLSP-----ACQNPLKKMVEEFVPR-----APRSTVLYLGTR 206
Db 333 E--MLNPITGTSYAAFEETLDYVTSK-IPRFPDFKFKGERBELGTQMKATGEWMAIGRT- 388
QY 207 GKHSLFEREIPEVLGLTFDHP-----GRMPDL-----ILHDEVRGWLFLMEAVKSKG 254
Db 389 -----YEESLKAINSLSYGVHHLGSLNGESVELDYIKERIGHQDDRLFFICEAIR--- 440
QY 255 PFDEHRSLQELFVTPSAGLIFVNCFN 283
Db 441 -----RGTSLEELHNWTKIDYFFLNKFN 464

RESULT 13
Q8E409
ID Q8E409 PRELIMINARY; PRT; 459 AA.
AC Q8E409;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN GBS1594.

OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RA MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RL "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766852; CAD47253.1; -
DR Sagallst; GBS1594; -
DR GO; GO:0003977; F:UDP-N-acetylglucosamine diphosphorylase act. .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR005882; GlmU.
DR InterPro; IPR001451; Hexapep transf.
DR InterPro; IPR005835; NTP transferase.
DR Pfam; PF00132; Hexapep; 7.
DR Pfam; PF00483; NTP transferase; 1.
DR TIGRFAMs; TIGR01173; GlmU; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 459 AA; 49377 MW; 4D6FE3E9E4629658 CRC64;

Query Match 6.0%; Score 100.5; DB 16; Length 459;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 60; Conservative 37; Mismatches 94; Indels 59; Gaps 13;

QY 16 TARALLKRFQDAQRVNVRSVATLLALAGLXPGDRWVDSTTPRLGVOKIMDWSGSHWAKP 75
Db 107 TGESLKNLIGHVNHKNVATILT-----ADANP-FGYGRIIRNSDEVTKI 152
QY 76 YATGSREDFRKKTLRQWVNGFAVLNADNL-----NIATNSQLNECYCLSD-----EALQ 124
Db 153 VEQKDANDFEQV--KEINTGTVYFDNQSLFEALKDINTNNAQGEYLTLDVIGIFKEAGK 210
QY 125 ALRAYGTGFPESL-----VFLDEASKAVK--ARAEL-----QAAMISVDLPGE 169
Db 211 KVGAYKLKDFDESIGVNDRAVALATAEKVMHRIARQHMVNGVTVVNPDSAYIDIDVEIGE 270
QY 170 EFLSLPAQNPPLKKMVEEFVPRFAPRSTVL-----YLGDTL-RGXSLPEREIFERVL--- 221
Db 271 ESIVPE---NVLTKGQTK-----IGKGTLLTNGSYLVDAQVGNVDVTTNSMVEESIISD 321
QY 222 GLTFDPHGMR 231
Db 322 GVTGVPYAH 331

RESULT 14
Q77372
ID Q77372 PRELIMINARY; PRT; 498 AA.
AC Q77372;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag polyprotein.
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94149849; PubMed=8107220;
RA Vanden Haesevelde M., Decourt J.L., De Leys R.J., Vanderborght B.,
RA van der Groen G., van Heuverswijn H., Saman E.;
RT "Genomic cloning and complete sequence analysis of a highly divergent
RT African human immunodeficiency virus isolate.";
RN J. Virol. 68:1586-1596(1994).
```

```

DR EMBL; L20587; AAA99878.1; -
DR HSPSP; P05888; IAA.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag p24.
DR InterPro; IPR000071; Retrovir p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS50158; zf_CCHC; 2.
KW AIDS; Core protein; Polyprotein.
SQ SEQUENCE 498 AA; 55304 MW; 64D16C9542519693 CRC64;

Query Match
Best Local Similarity 5.9%; Score 100; DB 15; Length 498;
Matches 65; Conservative 33; Mismatches 100; Indels 66; Gaps 14;

QY 48 GD---RWDDSTPLGLGVOKIMDSGEHWAKPYATGSREDFR---KKTLRQWVDNGFAVLN 101
Db 258 GDLYRKWI-----VLGLNKW-----KMYSPVSLIDIKQGPKEPRDYVDRFYKTLR 304

QY 102 ADNLNIAATNSQLNEYCL---SDEALQALRAYGTG-FEESLVVFLDEASKAVKARAEAL 156
Db 305 AEQATQEVKXWMTETLLVQVANPCKQLKSLGPGATLEENWVACQGVGGPTHKARVLAE 364

QY 157 QAAMISVDLPGGEEFLSPAGQNPLLKKWVEEF-----VPR--FAPRSTVLY----- 201
Db 365 AMATAQODLKGXYTAVPMQRGQNPIRKGTIKCFNCKGKEGHIAHRCAPRKKGCWKCGEG 424

QY 202 --LGDTK-GRKHSLFEREIFEIVELGLTFDPHGRMPDLI-----LHDEVKGLFLM 247
Db 425 HQMKDCENGQANF-----LGKYPPGCTREGNVQRPAPHSAPPMEEVKG----Q 472

QY 248 EAVKSK-GPFDEERHRSQLQELFVT 270
Db 473 ENQEQKGGPNELYPFASLKSFLGT 496

```

[illegible]

Search completed: October 1, 2004, 16:10:31  
Job time : 35.5414 secs

```

RESULT 15
Q06907 PRELIMINARY; PRT: 772 AA.
ID Q06907
AC Q06907;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE VP4 protein.
DE VP4.
OS Bovine rotavirus.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
NCBI_TaxID=10927;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=KK-3;
RC MEDLINE=94049850; PubMed=8232344;
RX Isegawa Y., Nakagomi O., Nakagomi T., Ishida S., Ueda S.;
RT "Determination of bovine rotavirus G and P serotypes by polymerase
chain reaction.";
RL Mol. Cell. Probes 7:277-284(1993).
DR ENBL; D14367; BAA03284.1; -.
DR CG; GO:0019028; C:viral capsid; IEA.
DR InterPro; IPR000416; Cap VP4.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR Pfam; PF000426; VP4; 1.
DR SQ SEQUENCE 772 AA; 86724 MW; E9A29D8B9EF9F15 CRC64;

```

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 42.3214 Seconds  
(without alignments)  
2156.424 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIQGTVAISIDTARAL.....EDPDHLHLNGSRFLGPIYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	7.0	708	4 AAB76666	Aab76666 Coryneb
2	118	7.0	713	4 AAB76665	Aab76665 Coryneb
3	118	7.0	1111	4 AAG92827	Aag92827 C glutami
4	105	6.2	135	6 ABM69452	Abm69452 Phototrab
5	103.5	6.1	1037	5 ABP39949	Abp39949 Staphyloc
6	103.5	6.1	1120	6 ABU42979	Abu42979 Protein e
7	100	5.9	772	6 AAU57632	Aau57632 VP4 prote
8	99.5	5.9	1057	6 ABU43956	Abu43956 Protein e
9	97.5	5.8	289	4 AAU36347	Aau36347 Pseudomon
10	97.5	5.8	289	6 ABU38597	Abu38597 Protein e
11	96.5	5.7	513	2 AAU5897	Aay5897 Vicia sat
12	95.5	5.7	737	4 AAU33575	Aau33575 Pseudomon
13	95.5	5.7	797	6 ABU15621	Abu15621 Protein e
14	95.5	5.7	1057	6 ABJ18981	Abj18981 Pathogen
15	95.5	5.7	1057	6 ABU42381	Abu42381 Protein e
16	95.5	5.7	1057	6 ABM71258	Abm71258 Staphyloc
17	94.5	5.6	574	4 ABBS68101	Abbs68101 Drosophil
18	94	5.6	155	4 AAG92225	Aag92225 C glutami
19	93.5	5.5	459	5 ABP26548	Abp26548 Streptoco
20	93	5.5	498	2 AAR51691	Aar51691 HIV-type
21	93	5.5	498	2 AAN33076	Aan33076 HIV isola
22	93	5.5	1080	4 AAM00803	Aam00803 Human bon
23	93	5.5	1637	4 AAM00916	Aam00916 Human bon
24	93	5.5	2263	4 AAM79000	Aam79000 Human pro
25	92.5	5.5	453	4 AAB96341	Aab96341 Putative

## ALIGNMENTS

### RESULT 1

AAB76666  
ID AAB76666 standard; protein; 708 AA.

XX AC

XX AAB76666;

XX AC

XX 11-APR-2001 (first entry)

XX XX

XX 11-APR-2001 (first entry)

XX XX

XX Corynebacterium glutamicum MCT protein SEQ ID NO:314.

XX KW

XX Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;

XX KW

XX membrane construction and membrane transport protein; petroleum spill;

XX KW

XX hydrocarbon degradation; gram positive aerobic bacterium; marker;

XX KW

XX identification; microorganism; fine chemical production; transformation;

XX KW

XX genome mapping; genetic engineering.

XX XX

XX Corynebacterium glutamicum.

XX OS

XX Corynebacterium glutamicum.

XX PN

XX WO200100805-A2.

XX XX

XX 04-JAN-2001.

XX PD

XX 23-JUN-2000; 2000WO-IB000926.

XX PF

XX 25-JUN-1999; 99US-0141031P.

XX PR

XX 08-JUL-1999; 99DE-01031454.

XX PR

XX 08-JUL-1999; 99DE-01031478.

XX PR

XX 08-JUL-1999; 99DE-01031563.

XX PR

XX 09-JUL-1999; 99DE-01032122.

XX PR

XX 09-JUL-1999; 99DE-01032124.

XX PR

XX 09-JUL-1999; 99DE-01032125.

XX PR

XX 09-JUL-1999; 99DE-01032128.

XX PR

XX 09-JUL-1999; 99DE-01032180.

XX PR

XX 09-JUL-1999; 99DE-01032182.

XX PR

XX 09-JUL-1999; 99DE-01032190.

XX PR

XX 09-JUL-1999; 99DE-01032191.

XX PR

XX 09-JUL-1999; 99DE-01032209.

XX PR

XX 09-JUL-1999; 99DE-01032212.

XX PR

XX 09-JUL-1999; 99DE-01032227.

XX PR

XX 09-JUL-1999; 99DE-01032228.

XX PR

XX 09-JUL-1999; 99DE-01032229.

XX PR

XX 09-JUL-1999; 99DE-01032230.

XX PR

XX 14-JUL-1999; 99DE-01032237.

XX PR

XX 14-JUL-1999; 99DE-01033005.

XX PR

XX 14-JUL-1999; 99DE-01033006.

XX PR

XX 27-AUG-1999; 99DE-01040764.

XX PR

XX 27-AUG-1999; 99DE-01040765.

XX PR

XX 27-AUG-1999; 99DE-01040766.

PR 27-AUG-1999; 99DE-01040830.  
 PR 27-AUG-1999; 99DE-01040831.  
 PR 27-AUG-1999; 99DE-01040832.  
 PR 27-AUG-1999; 99DE-01040833.  
 PR 31-AUG-1999; 99DE-01041378.  
 PR 31-AUG-1999; 99DE-01041379.  
 PR 31-AUG-1999; 99DE-01041395.  
 PR 03-SEP-1999; 99DE-01042077.  
 PR 03-SEP-1999; 99DE-01042078.  
 PR 03-SEP-1999; 99DE-01042079.  
 PR 03-SEP-1999; 99DE-01042088.  
 XX (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 XX WPI; 2001-071486/08.  
 DR N-PSDB; AAF67899.  
 XX  
 XX Corynebacterium glutamicum nucleic acids encoding membrane construction  
 PT and membrane transport proteins or their portions, useful for typing or  
 PT identifying C. glutamicum or related bacteria, and as markers for  
 PT transformation.  
 XX  
 PS Claim 20; Page 617-619; 1119pp; English.  
 XX  
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
 CC construction and membrane transport (MCT) proteins given in AAB76510 to  
 CC AAB76847. The MCT nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for modulating fine chemical production in C. glutamicum or  
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
 CC identification of C. glutamicum or related bacteria, as reference points  
 CC for mapping C. glutamicum genome, and as markers for transformation.  
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
 CC example from the present invention  
 XX  
 SQ Sequence 708 AA;  
 Query Match 7.0%; Score 118; DB 4; Length 708;  
 Best Local Similarity 25.5%; Pred. No. 0.026;  
 Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;  
 QY 29 QRYNRSVATLLALAGLPGDRWVSTTTRLCGVQKIMDSGE---HWAK-----P 75  
 Db 167 KFNIRCVTPPELSALGPRDGVTI----LGVRFQAQTGPTTRWEKDKRKLGRST 222  
 QY 76 YATGSREDFRKTLRQWDNGFAVLN-ADNLNIATNSQLNE----YCLSDALQALRAYG 130  
 Db 223 YRLGSTNDKAVETLRETQKAGKAVQADNRILAAANRAELRELERQYQASQELIKVSWAQI 282  
 QY 131 TEGFESLVVFLD-----EASKVKARAEALQAMISVDLPGGREF 171  
 Db 293 DVESAIAAEILDLLELNTPTEATLSARHEAAQTLLARVSDLLVAQSEETVASMNL 342  
 QY 172 LLSPAGQPLLKWMVEEFVPEFAPRSTVLVLDGTRGKSLPERFEFVLGTFD 226  
 Db 343 KRAETELKRLSLPVAVSEETAREVEKFLANTRVHA---ANVDEQTIALRED 394  
 RESULT 2  
 AAB76665  
 ID AAB76665 standard; protein; 713 AA.  
 XX  
 AC AAB76665;  
 XX  
 DT 11-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum MCT protein SEQ ID NO:312.  
 XX  
 KW Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT;  
 KW membrane construction and membrane transport protein; petroleum spill;  
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;

KW identification; microorganism; fine chemical production; transformation;  
 KW genome mapping; genetic engineering.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 FN WO200100805-A2.  
 XX  
 XX 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB000926.  
 XX  
 PR 25-JUN-1999; 99US-0141031P.  
 PR 08-JUL-1999; 99DE-01031454.  
 PR 08-JUL-1999; 99DE-01031478.  
 PR 08-JUL-1999; 99DE-01031563.  
 PR 09-JUL-1999; 99DE-01032122.  
 PR 09-JUL-1999; 99DE-01032124.  
 PR 09-JUL-1999; 99DE-01032125.  
 PR 09-JUL-1999; 99DE-01032128.  
 PR 09-JUL-1999; 99DE-01032180.  
 PR 09-JUL-1999; 99DE-01032182.  
 PR 09-JUL-1999; 99DE-01032190.  
 PR 09-JUL-1999; 99DE-01032191.  
 PR 09-JUL-1999; 99DE-01032209.  
 PR 09-JUL-1999; 99DE-01032212.  
 PR 09-JUL-1999; 99DE-01032227.  
 PR 09-JUL-1999; 99DE-01032228.  
 PR 09-JUL-1999; 99DE-01032229.  
 PR 09-JUL-1999; 99DE-01032230.  
 PR 14-JUL-1999; 99DE-01032927.  
 PR 14-JUL-1999; 99DE-01033005.  
 PR 14-JUL-1999; 99DE-01033006.  
 PR 27-AUG-1999; 99DE-01040764.  
 PR 27-AUG-1999; 99DE-01040765.  
 PR 27-AUG-1999; 99DE-01040766.  
 PR 27-AUG-1999; 99DE-01040830.  
 PR 27-AUG-1999; 99DE-01040831.  
 PR 27-AUG-1999; 99DE-01040832.  
 PR 27-AUG-1999; 99DE-01040833.  
 PR 31-AUG-1999; 99DE-01041378.  
 PR 31-AUG-1999; 99DE-01041379.  
 PR 31-AUG-1999; 99DE-01041395.  
 PR 03-SEP-1999; 99DE-01042077.  
 PR 03-SEP-1999; 99DE-01042078.  
 PR 03-SEP-1999; 99DE-01042079.  
 PR 03-SEP-1999; 99DE-01042088.  
 XX (BADI ) BASF AG.  
 PA  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 XX WPI; 2001-071486/08.  
 DR N-PSDB; AAF67898.  
 XX  
 XX Corynebacterium glutamicum nucleic acids encoding membrane construction  
 PT and membrane transport proteins or their portions, useful for typing or  
 PT identifying C. glutamicum or related bacteria, and as markers for  
 PT transformation.  
 XX  
 PS Claim 20; Page 612-614; 1119pp; English.  
 XX  
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
 CC construction and membrane transport (MCT) proteins given in AAB76510 to  
 CC AAB76847. The MCT nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for modulating fine chemical production in C. glutamicum or  
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
 CC identification of C. glutamicum or related bacteria, as reference points  
 CC for mapping C. glutamicum genome, and as markers for transformation.  
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
 CC example from the present invention  
 XX  
 SQ Sequence 713 AA;

Query Match 7.0%; Score 118; DB 4; Length 713;  
 Best Local Similarity 25.5%; Pred. No. 0.026;  
 Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNVRSAVTLALAGLPGDRWDSTTPRLGVQKIMDWSGE---HWAK-----P 75  
 DB 172 KRFNRCVTPPELSALGPRDQGVTI---LGVKFAQQTGPTTWEKDRRKLGDST 227  
 QY 76 YATGSRDFRKTTLRWQVNGFAVLN-ADNLMIATNSQLNE---YCLSDALQALRAYG 130  
 DB 228 YRLGSTNDKAVETLRETQKAGKAVVQADNRILAAARAELELERQYQASQETILKVSQAQI 287  
 QY 131 TGGFESLVVFLD-----EASKAVKARAEALQAAAMISVDLPQGEF 171  
 DB 288 DYESADAATAELDRLEELNTPTEATLSARHEAAKQTLARVSDLLVAASQSETVASMM 347  
 QY 172 LLSPAGQNPLKKWVEEFVFPFAPRSTVLYLGDTRGKHSLFEREIPEEVLGLTFD 226  
 DB 348 KRAETELKRLSLPVAEVSEETAREVEKLFANTRRVHA---ANVDEQITIALRED 399

RESULT 3  
 AAG92827  
 ID AAG92827 standard; protein; 1111 AA.  
 AC AAG92827;  
 DT 26-SEP-2001 (first entry)  
 DE C glutamicum protein fragment SEQ ID NO: 6581.  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 OS Corynebacterium glutamicum.  
 XX EP1108790-A2.  
 XX 20-JUN-2001.  
 XX 18-DEC-2000; 2000EP-00127688.  
 XX 16-DEC-1999; 99JP-00377484.  
 XX 07-APR-2000; 2000JP-00159162.  
 XX 03-AUG-2000; 2000JP-00280986.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;  
 XX Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI; 2001-376931/40.  
 XX N-PSDB; AAH68046.  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analyzing  
 PT expression profile or pattern of a gene and identifying homologous gene.  
 XX Claim 17; SEQ ID NO 6581; 246pp + Sequence Listing; English.  
 XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and analysing  
 CC the expression profile or expression pattern of a gene derived from  
 CC Coryneform bacterium, and identifying a homologue of a gene derived from  
 CC Coryneform bacterium. Coryneform bacteria are useful for producing amino  
 CC acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described in the  
 CC exemplification of the invention. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from the European Patent Office

XX SQ Sequence 1111 AA;  
 Query Match 7.0%; Score 118; DB 4; Length 1111;  
 Best Local Similarity 25.5%; Pred. No. 0.052;  
 Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNVRSAVTLALAGLPGDRWDSTTPRLGVQKIMDWSGE---HWAK-----P 75  
 DB 570 KRFNRCVTPPELSALGPRDQGVTI---LGVKFAQQTGPTTWEKDRRKLGDST 625  
 QY 76 YATGSRDFRKTTLRWQVNGFAVLN-ADNLMIATNSQLNE---YCLSDALQALRAYG 130  
 DB 626 YRLGSTNDKAVETLRETQKAGKAVVQADNRILAAARAELELERQYQASQETILKVSQAQI 685  
 QY 131 TGGFESLVVFLD-----EASKAVKARAEALQAAAMISVDLPQGEF 171  
 DB 686 DYESADAATAELDRLEELNTPTEATLSARHEAAKQTLARVSDLLVAASQSETVASMM 745  
 QY 172 LLSPAGQNPLKKWVEEFVFPFAPRSTVLYLGDTRGKHSLFEREIPEEVLGLTFD 226  
 DB 746 KRAETELKRLSLPVAEVSEETAREVEKLFANTRRVHA---ANVDEQITIALRED 797

RESULT 4  
 ABM69452  
 ID ABM69452 standard; protein; 135 AA.  
 XX ABM69452;  
 AC ABM69452;  
 DT 20-NOV-2003 (first entry)  
 DE Photorhabdus luminescens protein sequence #2549.  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX Photorhabdus luminescens.  
 XX WO200294867-A2.  
 XX 28-NOV-2002.  
 XX 07-FEB-2002; 2002WO-IB003040.  
 XX 07-FEB-2001; 2001FR-00001659.  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 XX Buchrieser C;  
 XX WPI; 2003-148459/14.  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX Claim 2; SEQ ID NO 2549; 1205pp; French.  
 XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.





polypeptide or its fragment whose expression is inhibited by the  
 antisense nucleic acid; (4) an antibody capable of specifically binding  
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 proliferation or the activity of a gene in an operon required for  
 proliferation; (7) identifying a compound that influences the activity of  
 the gene product or that has an activity against a biological pathway  
 required for proliferation, or that inhibits cellular proliferation; (8)  
 identifying a gene required for cellular proliferation or the biological  
 pathway in which a proliferation-required gene or its gene product lies  
 or a gene on which the test compound that inhibits proliferation of an  
 organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 compound's activity; (11) a culture comprising strains in which the gene  
 product is overexpressed or underexpressed; (12) determining the extent  
 to which each of the strains is present in a culture or collection of  
 strains; or (13) identifying the target of a compound that inhibits the  
 proliferation of an organism. The antisense nucleic acids are useful for  
 identifying proteins or screening for homologous nucleic acids required  
 for cellular proliferation to isolate candidate molecules for rational  
 drug discovery programs, or for screening homologous nucleic acids  
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
*K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 the target prokaryotic essential genes. Note: The sequence data for this  
 patent did not form part of the printed specification, but was obtained  
 in electronic format directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1120 AA;

Query Match 6.1%; Score 103.5; DB 6; Length 1120;  
 Best Local Similarity 22.7%; Pred. No. 1.5;  
 Matches 61; Conservative 41; Mismatches 82; Indels 85; Gaps 14;  
 QY 82 EDPRKKTLEWDVNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
 Db 573 KELEYEVMKDNDAIVNCNMENIDPVGHTGDSIVVAPSQTLSDEVYQMLRDVSLKVR 632  
 QY 128 AXYTEG-----FESLVVFLDEASKAKARAEALQ-----AAMISVDLPGG 168  
 Db 633 ALGIEGCVQLALDPSLNYIIEVNPV--SHSSALASKATGYPIAKLAAGVGLTD 691  
 QY 169 EEFLLSP-----AGQNPLLLKQWEEFVPRF-----APRSTVLYLGDT 206  
 Db 692 E--MLNPITGTSYAAPEPTLDYVISK-IPRFPDFEKGGERELGTQMKATGEVMAIGRT- 747  
 QY 207 GKHSLEFPERIFEVLGLTFD-----PHGRMPDL-----ILHDEVGRWMLFMEAVKSKG 254  
 Db 748 -----YEESLLKAIRSLFVGVHGLPNGESYELDYIKERIGHQDDERLFFIGEAIR--- 799  
 QY 255 PFDEERHRSLOELFVTPSAGLIFVNCFN 283  
 Db 800 -----RGTSLEELHNTKIDYFLNKFQN 823

RESULT 7  
 AAW57632  
 ID AAW57632 standard; protein; 772 AA.  
 AC AAW57632;  
 XX  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 27-AUG-1998 (first entry)  
 XX  
 XX  
 DE VP4 protein of human rotavirus isolate 116E.  
 XX  
 KW VP7 gene; human rotavirus; rotavirus strain G9P11; bovine rotavirus;  
 KW VP4 gene; mucosal immunity stimulation; infection.  
 XX  
 OS Human rotavirus A.  
 XX  
 PN US5773009-A.  
 XX  
 PD 30-JUN-1998.  
 XX

19-FEB-1997; 97US-00802141.  
 15-APR-1994; 94US-00231041.  
 (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 Bhan MK, Das BK, Glass RI, Gentsch JR;  
 WPI; 1998-386933/33.  
 N-PSDB; AAV24477.  
 New isolated rotavirus strain G9P11 - used for stimulating immunity  
 against infection by virulent strains of rotaviruses.  
 Example 3; Col 33-38; 24pp; English.  
 This sequence is the VP4 protein of human rotavirus strain 116E. The DNA  
 sequence can be used in the rotavirus of the invention which is an  
 isolated rotavirus strain G9P11 comprising a bovine viral VP4 virus gene  
 and a human VP7 virus gene. The new avirulent rotavirus strain G9P11 can  
 be used for stimulating mucosal immunity against infection by virulent  
 strains of rotaviruses, e.g. strain serotype G10P11. Rotaviruses are  
 responsible for diarrhoea in infants, especially in developing countries.  
 It can also be used for producing antigens and antibodies which can be  
 used for diagnosis. (Updated on 17-OCT-2003 to standardise OS field)  
 Sequence 772 AA;  
 Query Match 5.9%; Score 100; DB 2; Length 772;  
 Best Local Similarity 22.6%; Pred. No. 1.8;  
 Matches 76; Conservative 48; Mismatches 128; Indels 84; Gaps 16;  
 QY 1 VNSSDGDIGTASIDTARALLKRFDFDAQRYNVRSAVTLALA-----GLKPGDRWVDS 55  
 Db 473 VPSNDYQTPIANSVTVRQDLER-QLDENRREFNELSANIALSQLIDLALLP-----LDMF 527  
 QY 56 TPLRGVQKIMDSGEHWAKPYATGSRDPRK-----KTLRQ 91  
 Db 528 SMFSGTRSTIE-----AAKNFATSVMKFRKSNLAKSVNSLTDITDAAGSISRSTLS 582  
 QY 92 -----WVDNGFAVLNADNLNATNSQLNNEYCLSDALQALRAYGTEGF--EESLVVFL 142  
 Db 583 ANSAVSVMTDISDIVSDTNVVTAT-----ATAAAKFRVKEFTTEFGVSD 630  
 QY 143 DEASKAKVAKARAEALQAAAMISVDLPGGEEFLSPAGQNPLLKQWEEFVPRFAPR---STV 199  
 Db 631 DISAAVVKTKMKNLV---VD---EEML-----PQITTEASEKFIENRAYRLIDGDK 676  
 QY 200 LYLGDTRGKHSLEFPERIFEVLGLTFDPHGRMPDLILHDEVGRWMLFMEAVKS-KGPFDE 258  
 Db 677 VYEVTEGKYFAYLTETFEVM---FDAE-RAELVTVSPVISAIIDFKIKLNNDNYGI 732  
 QY 259 ERHRSLOELFVTPSAGLIFVNCFNRESMRQWLPEL 294  
 Db 733 TREQALNMLRSDPKVLRSFIN--QNNPIKNRIEQL 766  
 RESULT 8  
 ABU43956  
 ID ABU43956 standard; protein; 1057 AA.  
 XX  
 AC ABU43956;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #29483.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Staphylococcus haemolyticus.  
 XX  
 PN WO200277183-A2.  
 XX

```
PD 03-OCT-2002.
XX
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang I, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX N-PSDB; ACA47826.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 71880; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1057 AA;
XX
XX Query March 5.9%; Score 99.5; DB 6; Length 1057;
XX Best Local Similarity 21.9%; Pred. No. 3.4;
XX Matches 59; Conservative 42; Mismatches 83; Indels 85; Gaps 14;
XX
XX 82 EDRKKTUQWVDNGFAVINDNLN-----IATNSQLN--EY-CLSDALQALR 127
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 214 KEIEVVRKNDNAIVVCNMENIDPVCIHTGDSIVAPSTQSLSDVEQMLRVSCLKVIR 273
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 128 AYTGET-----FEESLVVFLDEASKAVKARALQ-----AMISVDLPGG 168
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 274 ALGIEGCGNQLALDPHSFNVIIEVNPV--SRSSALASKATGYBIAKLAIAVGLTLD 332
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 169 EEFILSP-----AQONPLKKWVEEFVPRF-----APRSTVLVLDTR 206
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 333 E--MLNPVTGTSYAAFEPLTDYVISK--IPRPFPDFKEKGERELGTQMKATGEVMAIGT-- 388
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX 207 GKHSLFEREIFEYVLGLTFD-----PHGRMPDL-----ILHDEVRGWLFLEAVKSKG 254
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 389 -----YEESLLKAIRSLEYGVVHHLGLNGESFDLDYIKERISHQDDERLFFIGAIR--- 440
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 255 PFDEERHRSQELFPVTPSAGLIFVNCFEN 283
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 441 -----RGTLEEIHNTQIDYFFLNKFN 464
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX
XX RESULT 9
XX AAU36347
XX ID AAU36347 standard; protein; 289 AA.
XX
XX AC AAU36347;
XX
XX DT 14-FEB-2002 (first entry)
XX
XX DE Pseudomonas aeruginosa cellular proliferation protein #337.
XX
XX KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
XX antibacterial; drug design.
XX
XX OS Pseudomonas aeruginosa.
XX
XX PN WO200170955-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 21-MAR-2001; 2001WO-US009180.
XX
XX PR 21-MAR-2000; 2000US-0191078P.
XX 23-MAY-2000; 2000US-0206848P.
XX 26-MAY-2000; 2000US-0207727P.
XX 23-OCT-2000; 2000US-0242578P.
XX 27-NOV-2000; 2000US-0253625P.
XX 22-DEC-2000; 2000US-0257931P.
XX 16-FEB-2001; 2001US-0269308P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX DR WPI; 2001-611495/70.
XX N-PSDB; AAS54206.
XX
XX DR New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Example 3; SEQ ID NO 11940; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes,
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence represents an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 289 AA;
```

Query Match 5.8%; Score 97.5; DB 4; Length 289;  
Best Local Similarity 24.0%; Pred. No. 0.72;  
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;

QY 94 DNGFAVLNADNLNATNSQLNEYCLSDALQALRAYGTGFEESLVVFLDEASKAVKARA 153  
Db 69 DNKAAVI-----IEVNSQDFFLALQDD-----FKGFVAESLEKAFNEKLTDAAPLVEARE 118

QY 154 EALQAAM-----ISVDLPGBEFLLSPAGQNPLLK 183  
Db 119 EARLALVAKTGENVNIRRLTRVEGDVVGAYLHGHRIGVVNLKGG-----NPELA 168

QY 184 KMVEEFV-----PRFAPRSTVLYLGDTGRKHSLFERIFEVLGLTFDPHGRMPDLILHDE 239  
Db 169 KDIAHVAASNPOFLSASEVSEEAIAK-----EKEIF---LALNADKIAGKPNIVENM 219

QY 240 VRGWL--FLMEAVKSGPFDEERHRSLOELFVTPSAGLIFVNCFFENRESMRQWLPDLAW 297  
Db 220 VKGRISKFLAEASLVLEQFPFVKNPEVKVGLAKQAGABIVSFVRYEGEGIEKAEVDFAAE 279

QY 298 TEAWVA 303  
Db 280 VAAQVA 285

RESULT 10  
ABU38597  
ID ABU38597 standard; protein; 289 AA.  
AC ABU38597;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #24124.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Irawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA42467.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 66521; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding

the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
proliferation or the activity of a gene in an operon required for  
proliferation; (7) identifying a compound that influences the activity of  
the gene product or that has an activity against a biological pathway  
required for proliferation, or that inhibits cellular proliferation; (8)  
identifying a gene required for cellular proliferation or the biological  
pathway in which a proliferation-required gene or its gene product lies  
or a gene on which the test compound that inhibits proliferation of an  
organism acts; (9) manufacturing an antibiotic; (10) profiling a  
compound's activity; (11) a culture comprising strains in which the gene  
product is overexpressed or underexpressed; (12) determining the extent  
to which each of the strains is present in a culture or collection of  
strains; or (13) identifying the target of a compound that inhibits the  
proliferation of an organism. The antisense nucleic acids are useful for  
identifying proteins or screening for homologous nucleic acids required  
for cellular proliferation to isolate candidate molecules for rational  
drug discovery programs, or for screening homologous nucleic acids  
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
*K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
the target prokaryotic essential genes. Note: The sequence data for this  
patent did not form part of the printed specification, but was obtained  
in electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 289 AA;  
XX

Query Match 5.8%; Score 97.5; DB 6; Length 289;  
Best Local Similarity 24.0%; Pred. No. 0.72;  
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;

QY 94 DNGFAVLNADNLNATNSQLNEYCLSDALQALRAYGTGFEESLVVFLDEASKAVKARA 153  
Db 69 DNKAAVI-----IEVNSQDFFLALQDD-----FKGFVAESLEKAFNEKLTDAAPLVEARE 118

QY 154 EALQAAM-----ISVDLPGBEFLLSPAGQNPLLK 183  
Db 119 EARLALVAKTGENVNIRRLTRVEGDVVGAYLHGHRIGVVNLKGG-----NPELA 168

QY 184 KMVEEFV-----PRFAPRSTVLYLGDTGRKHSLFERIFEVLGLTFDPHGRMPDLILHDE 239  
Db 169 KDIAHVAASNPOFLSASEVSEEAIAK-----EKEIF---LALNADKIAGKPNIVENM 219

QY 240 VRGWL--FLMEAVKSGPFDEERHRSLOELFVTPSAGLIFVNCFFENRESMRQWLPDLAW 297  
Db 220 VKGRISKFLAEASLVLEQFPFVKNPEVKVGLAKQAGABIVSFVRYEGEGIEKAEVDFAAE 279

QY 298 TEAWVA 303  
Db 280 VAAQVA 285

RESULT 11  
AAY05897  
ID AAY05897 standard; protein; 513 AA.  
XX  
AC AAY05897;  
XX  
XX 02-AUG-1999 (first entry)  
XX  
DT Vicia sativa omega-myristic acid hydroxylase CYP94A2.  
DE  
DE Fatty acid omega-hydroxylase; omega-myristic acid hydroxylase;  
XX cytochrome P450; transgenic plant; lipid; hydroxylation; epoxidation;  
KW oilseed; vegetable oil; crop protection; omega-hydroxy acid; CYP94A2.  
XX  
XX Vicia sativa.  
XX  
XX Location/Qualifiers  
FH Key 319. .329  
FT Domain /note= "haem-binding domain, corresponds to motif claimed  
FT in Claim 2"  
XX  
PN WO9918224-A1.

XX PD 15-APR-1999.

XX PF 06-OCT-1998; 98WO-1B001716.

XX PR 06-OCT-1997; 97US-0060960P.

XX PA (CNRS ) CENT NAT RECH SCI.

XX PI Tijet N, Pinot F, Benveniste I, Le Bouquin R, Helvig C, Batard Y;

XX PI Cabello-Huatao F, Werck-Reichhart D, Salaun J, Durst F;

XX DR WPI; 1999-264030/22.

XX DR N-PSDB; AAX58401.

XX PT Nucleic acid encoding plant fatty acid hydroxylases.

XX PS Example 2; Fig 7; 157pp; English.

XX CC The present sequence represents *Vicia sativa* CYP94A2 omega-myristic acid

XX CC hydroxylase, a microsomal cytochrome P450-dependent hydroxylase which

XX CC catalyses the transformation of myristic acid into 14-

XX CC hydroxytetradecanoic acid (i.e. terminal methyl hydroxylation). Low

XX CC levels of transformation of lauric and palmitic acids into the

XX CC corresponding omega-hydroxy fatty acids are also observed. The invention

XX CC provides isolated nucleic acids (see AAX58400-06) encoding plant fatty

XX CC acid hydroxylases (see AAY05896-902). Also claimed are host cells,

XX CC transgenic plants and compositions consisting of the plant fatty acid

XX CC hydroxylase, a process for isolating additional fatty acid hydroxylase

XX CC genes from a plant, and a process of altering fatty acid composition in a

XX CC plant by expressing the plant fatty acid hydroxylase in a transgenic

XX CC plant, and hydroxylating or epoxidating a fatty acid substrate in the

XX CC plant. Manipulating the hydroxylated fatty acid content of plants will

XX CC modify resistance to drought and attack by insects and other pests. The

XX CC transgenic plants may also be used as sources of hydroxylated and

XX CC epoxidized fatty acids useful in the manufacture of e.g. lubricants, anti

XX CC -slip agents, plasticisers, coating agents, detergents and surfactants

XX CC

XX SQ Sequence 513 AA;

Query Match 5.7%; Score 96.5; DB 2; Length 513;

Best Local Similarity 22.6%; Pred. No. 2.2;

Matches 72; Conservative 44; Mismatches 112; Indels 91; Gaps 15;

QY 64 IMDWSGEHWKPYATGSRDRKTKLTROWVNGFAVLNADN----LNATNSQLN-----114

Db 126 IFNADGESWKFQRTISSHE-FNTRSLRKFVETVVDVLSLDRPLVPLVLSQANSQTLLDFQD 184

QY 115 -----EYC-----LSDEALQAL-----RAYG 130

Db 185 ILQRTFONICMIARGYPDEVLPSLPFPAKAFDESSQLSIERLNALIPLLWKVKRFL 244

QY 131 TEGFEESLVWFDE----ASKAVK-----ARAEALQAAIMI SVDLPGGEEFLLSAGQNP 181

Db 245 NIGVERQLKEAAVAVRGLATKIVKNKKELKEALQSESEVDLL--SRFLSSGHSDES 302

QY 182 LKKWVEEFVPPAPRST----VLYLGDTRGKHSLEPERIFEVLGLTDPHGMEDLLH 237

Db 303 VTDMMVISII--LAGRDTTSAALTWFFLLSKHSHVENEILKEITGKS-----ETGY 352

QY 238 DEVRGWLF----LMEAVSKGPFDEERHRSLOELFVTPSAGI-----IFVNCFFNRISM 287

Db 353 DEVKDMVYTHALCESMRLLPPLPVDTKVAVHD-DVLPDGLTKKGRVYTHIYAMGRSE 411

QY 288 ROWLPPELA-WETEAMVAED 305

Db 412 KIWGPDWAEFRPERWLSRD 430

RESULT 12

AAU33575

ID AAU33575 standard; protein; 797 AA.

XX

AC AAU33575;

XX DT 14-FEB-2002 (first entry)

XX DE Pseudomonas aeruginosa cellular proliferation protein #19.

XX DE DE Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX KW antibacterial; drug design.

XX KW Pseudomonas aeruginosa.

XX OS WO200170955-A2.

XX PN 27-SEP-2001.

XX PD 21-MAR-2001; 2001WO-US009180.

XX PF 21-MAR-2000; 2000US-0191078P.

XX PR 23-MAY-2000; 2000US-0206848P.

XX PR 26-MAY-2000; 2000US-0207727P.

XX PR 23-OCT-2000; 2000US-0242578P.

XX PR 27-NOV-2000; 2000US-0253625P.

XX PR 22-DEC-2000; 2000US-0257931P.

XX PR 16-FEB-2001; 2001US-0269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX XX WPI; 2001-611495/70.

XX DR N-PSDB; AAS51434.

XX XX New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids.

XX XX

XX PS Example 3; SEQ ID NO 5071; 511pp; English.

XX CC The invention relates to antisense inhibitors of genes essential to

XX CC prokaryotic cellular proliferation, their use in identifying the genes,

XX CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*

XX CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, also

XX CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also

XX CC useful for the identification of potential new targets for antibiotic

XX CC development. The antisense nucleic acids can also be used to identify

XX CC proteins used in proliferation, to express these proteins, and to obtain

XX CC antibodies capable of binding to the expressed proteins. The proteins can

XX CC be used to screen compounds in rational drug discovery programmes. The

XX CC antisense nucleic acid sequence is also useful to screen for homologous

XX CC nucleic acids which are required for cell proliferation in a wide variety

XX CC of organisms. The present sequence represents an essential prokaryotic

XX CC cellular proliferation protein. Note: The sequence data for this patent

XX CC did not form part of the printed specification, but was obtained in

XX CC electronic format directly from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 797 AA;

Query Match 5.7%; Score 95.5; DB 4; Length 797;

Best Local Similarity 24.7%; Pred. No. 5.5;

Matches 80; Conservative 36; Mismatches 91; Indels 117; Gaps 22;

QY 10 TVASIDTARALLKRGFDQAQ-RYNVRSAVTLLA-----LAGLXPG---DRWVDST 55

Db 325 TQFSIDNSTLGLWYNWDSRVYANRAEQMLGHADGQLVDRPLADFEFGGLMDRWLNL- 383

QY 56 TPRLGQVKIMDWSGEHWKPYATGSR-----DPRKTKLTROWVNGFAVLNADNLTATNSQ 112

Db 384 -----WRR--ARNSEEGPLSPETRCR-----AD-----405

QY 113 LNEYCLSDALQALRAYGTGTEGFEESLVVFLDEASKAVKARAEALQ---AAM--ISVDLP 167

XX

Db 406 -GSMPLPADVLSFLR-FGT-----SEYLVVFLSDVTTERRAR-EALQSEARMKGIASNVPG 459

QY 168 GEEFLLSPAGCNLLKKMVEFV-PRAPRSTVLYLGDTRGKHSIF---EREIFEVLGL 223

Db 460 -----MVFLEPRAGATSDPAYISE--GSEALVGYSAKELIESGRGI 500

QY 224 TFDPHGRMPDILLHDEVRGWLFLMEAVKSGPPDEERHRSIQELFVTPSAGLIFVN---- 279

Db 501 RGLVH---PD-----DRERYWSSQMAAL-----DENRDWHQGRILTRQGLRWADIKAS 547

QY 280 --CFENRESMRQWLPALAWETEAW 301

Db 548 ARCFEDGRAV-----WDGVVW 563

RESULT 13

ABU15621

ID ABU15621 standard; protein; 797 AA.

XX AC ABU15621;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #1148.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Pseudomonas aeruginosa.

XX PN WO20027183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.

XX DR N-PSDB; ACA19491.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 43545; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 797 AA;

Query Match 5.7%; Score 95.5; DB 6; Length 797;

Best Local Similarity 24.7%; Pred. No. 5.5;

Matches 80; Conservative 36; Mismatches 91; Indels 117; Gaps 22;

QY 10 TVASIDTARALLKRFQFDAQ-RYNVRSANVTLLA-----LAGLKPQ---DRWVDST 55

Db 325 TQFSIDNSTIGILVWVMDSRVYANRAAEQMLGHADGQLVDRPLADPEGLDMRWLNL- 383

QY 56 TPLGVQKIMDWSGEHWAKPYATGSRE---DERKKTILROWVDNGFAVLNADNLNIATNSQ 112

Db 384 -----WRR--ARNSEEGPLSFETRCLR-----AD----- 405

QY 113 LNEYCLSDAQLRAYGTGTFEESLVVFLDEASKAVKARAEALQ---AAM--ISVDLPQ 167

Db 406 -GSMPLPADVLSFLR-FGT---SEYLVVFLSDVTTERRAR-EALQSEARMKGIASNVPG 459

QY 168 GEEFLLSPAGCNLLKKMVEFV-PRAPRSTVLYLGDTRGKHSIF---EREIFEVLGL 223

Db 460 -----MVFLEPRAGATSDPAYISE--GSEALVGYSAKELIESGRGI 500

QY 224 TFDPHGRMPDILLHDEVRGWLFLMEAVKSGPPDEERHRSIQELFVTPSAGLIFVN---- 279

Db 501 RGLVH---PD-----DRERYWSSQMAAL-----DENRDWHQGRILTRQGLRWADIKAS 547

QY 280 --CFENRESMRQWLPALAWETEAW 301

Db 548 ARCFEDGRAV-----WDGVVW 563

RESULT 14

ABU18981

ID ABU18981 standard; protein; 1057 AA.

XX AC ABU18981;

XX DT 06-MAR-2003 (first entry)

XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 152.

XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;

XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;

XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;

XX KW autoimmune disease; HIV; hepatitis.

XX OS Staphylococcus sp.

XX PN WO200259148-A2.

XX PD 01-AUG-2002.

XX PF 21-JAN-2002; 2002WO-EP000546.

XX PR 26-JAN-2001; 2001AT-00000130.

XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.

XX PI Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;

PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;  
PI Tempelmaier B;  
XX WPI; 2003-075410/07.  
XX Identifying, isolating and producing hyperimmune serum-reactive antigens  
XX from a pathogen, for preparing vaccine or medicament for treating or  
XX preventing e.g. staphylococcal infections, comprises providing antibody  
XX preparation.  
XX Claim 21; Page 173; 252pp; English.  
XX The invention relates to a novel method for identifying, isolating and  
XX producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
XX allergen, a tissue or host prone to auto-immunity, where the antigens are  
XX used in a vaccine, comprises providing antibody preparation from a plasma  
XX pool of a type of animal, or individual sera with antibodies against the  
XX specific pathogen, tumour, allergen, tissue or host prone to auto-  
XX immunity. The hyperimmune serum-reactive antigens comprising any of the  
XX 62 sequences of 53-2261 amino acids fully defined in the specification,  
XX or their hyperimmune fragments are useful for the manufacture of a  
XX pharmaceutical preparation, particularly a vaccine against staphylococcal  
XX infections or colonisation against S. aureus or S. epidermidis. The  
XX preparation of antibodies is useful for the manufacture of a medicament  
XX for treating or preventing staphylococcal infections or colonisation  
XX against S. aureus or S. epidermidis. The antibody preparations may also  
XX be used for diagnostic and imaging purposes. Other conditions that can be  
XX treated include cancer, autoimmune diseases or infections caused by viral  
XX (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This  
XX sequence represents a staphylococcal protein relating to the method for  
XX identifying and producing pathogen specific antigens of the invention  
XX  
XX Sequence 1057 AA;  
Query Match 5.7%; Score 95.5; DB 6; Length 1057;  
Best Local Similarity 21.5%; Pred. No. 8.5;  
Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;  
QY 82 EDFRKKTLROWVDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
Db 214 KEIEYVNRKNDNAIVVNCNMENIDPVGHTGDSIVWAPSTLSVDEYQMLRDSLKVR 273  
QY 128 AYGTGEG-----FEESLVVFLDEASKAVKARAEALQ-----AAMISVDLPFG 168  
Db 274 ALGIEGGCNVQLALDPHSPDYVYIIIEVNPV-SRSSALASKATGYPIAKLAIAVGLTLD 332  
QY 169 EEFLSP-----AGQNPLKKXVVEFVPRF-----APRSTVLYLGDT 206  
Db 333 E--MLNPITGTSYAAFEPTLDYVISK-IPRFPFKPKGERELGTQMKATGEVMAIGRT- 388  
QY 207 GKHSIFEREIPEEVGLTFD-----PHGRMPDL-----ILHDEVRGWLFLMEAVKSG 254  
Db 389 -----YESLLKAIRSLSEYGVHHLGPNGESFDLDYIKERISHQDDERLFFGEAIR--- 440  
QY 255 PFDEERHRSLOSLFVTPSAGLIFVNCFNRESMROWLPE 293  
Db 441 -----RGVILEIHNTQIDYFLFKFQNIIDIEHQLKE 474  
RESULT 15  
ABU42381  
ID ABU42381 standard; protein; 1057 AA.  
XX ABU42381;  
AC ABU42381;  
XX 19-JUN-2003 (first entry)  
XX Protein encoded by Prokaryotic essential gene #27908.  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Staphylococcus aureus.  
XX

PN WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-USO09107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA46251.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 70305; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than S. aureus, S. typhimurium,  
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
XX the target prokaryotic essential genes. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1057 AA;  
Query Match 5.7%; Score 95.5; DB 6; Length 1057;  
Best Local Similarity 21.5%; Pred. No. 8.5;  
Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;  
QY 82 EDFRKKTLROWVDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
Db 214 KEIEYVNRKNDNAIVVNCNMENIDPVGHTGDSIVWAPSTLSVDEYQMLRDSLKVR 273  
QY 128 AYGTGEG-----FEESLVVFLDEASKAVKARAEALQ-----AAMISVDLPFG 168  
Db 274 ALGIEGGCNVQLALDPHSPDYVYIIIEVNPV-SRSSALASKATGYPIAKLAIAVGLTLD 332  
QY 169 EEFLSP-----AGQNPLKKXVVEFVPRF-----APRSTVLYLGDT 206

Db 333 E--MLNPTTCTSYAAFEFTLDYVISK-IPRFPDKFKGERELGTOMKATGEVMAIGRT- 388  
Qy 207 GKHSLPEREIFEVLGTED-----PHGRMPDL-----ILHDEVRCWLFLEAVKSKG 254  
Db 389 -----YEESLLKAIRSLEYGVHHLGLFNGESFDLDYKERISHODDERLFFIGEAIR--- 440  
Qy 255 PFDEERHRSLOELFVTPESAGLIFVNCFENRESMRQWLPE 293  
Db 441 -----RGTTLEEIHMTQIDYFFFLHKKFQNIIDIEHQLKE 474

Search completed: October 1, 2004, 16:13:00  
Job time : 47.3214 secs



Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:09:01 ; Search time 13.1755 Seconds  
(without alignments)  
1265.619 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDQIGTVASIDTARAL.....EDPDHLHLNGSRFLQGYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*
- 6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103.5	6.1	1037	US-09-134-001C-4794	Sequence 4794, Ap
2	100	5.9	772	US-08-802-141-4	Sequence 4, Appli
3	97.5	5.8	310	US-09-252-991A-21111	Sequence 21111, A
4	96	5.7	511	US-08-676-444-42	Sequence 42, Appl
5	95.5	5.7	835	US-09-252-991A-17364	Sequence 17364, A
6	93	5.5	498	US-08-470-202-59	Sequence 59, Appl
7	93	5.5	498	US-08-471-770-60	Sequence 59, Appl
8	93	5.5	498	US-08-468-059-59	Sequence 59, Appl
9	93	5.5	498	US-09-109-916-59	Sequence 59, Appl
10	93	5.5	498	US-09-886-156-59	Sequence 59, Appl
11	93	5.5	498	US-09-886-149-59	Sequence 59, Appl
12	93	5.5	498	US-09-886-150-59	Sequence 59, Appl
13	93	5.5	498	US-09-886-159-59	Sequence 59, Appl
14	92	5.4	171	US-09-540-236-2189	Sequence 22, Appl
15	91.5	5.4	1360	US-09-788-657-22	Sequence 4, Appli
16	91	5.4	506	US-09-299-662-4	Sequence 1, Appli
17	91	5.4	524	US-09-299-662-1	Sequence 66, Appl
18	90.5	5.4	503	US-09-408-020-66	Sequence 4, Appli
19	90	5.3	372	US-09-973-963-3	Sequence 4, Appli
20	88.5	5.2	501	US-09-489-039A-12663	Sequence 4, Appli
21	86.5	5.1	561	US-09-252-991A-16726	Sequence 12663, A
22	86	5.1	466	US-09-252-991A-32781	Sequence 16726, A
23	86	5.1	3472	US-09-408-020-4	Sequence 32781, A
24	84	5.0	450	US-09-134-001C-4811	Sequence 4, Appli
25	84	5.0	822	US-09-886-319A-64	Sequence 4811, Ap
26	83.5	4.9	877	US-08-208-036-14	Sequence 64, Appl
27	83.5	4.9	877	US-08-428-823-14	Sequence 14, Appl

28	83	4.9	295	4	US-09-252-991A-33101	Sequence 33101, A
29	83	4.9	348	3	US-08-855-910-8	Sequence 8, Appli
30	83	4.9	498	1	US-08-470-202-60	Sequence 60, Appl
31	83	4.9	498	1	US-08-471-770-60	Sequence 60, Appl
32	83	4.9	498	2	US-08-468-059-60	Sequence 60, Appl
33	83	4.9	498	3	US-09-109-916-60	Sequence 60, Appl
34	83	4.9	498	4	US-09-886-156-60	Sequence 60, Appl
35	83	4.9	498	4	US-09-886-149-60	Sequence 60, Appl
36	83	4.9	498	4	US-09-886-150-60	Sequence 60, Appl
37	83	4.9	498	4	US-09-886-159-60	Sequence 60, Appl
38	82.5	4.9	876	2	US-08-633-476-2	Sequence 2, Appli
39	82	4.9	223	4	US-09-489-039A-12066	Sequence 12066, A
40	82	4.9	893	4	US-09-514-302-4	Sequence 4, Appli
41	82	4.9	1938	4	US-09-514-302-2	Sequence 2, Appli
42	81.5	4.8	320	4	US-09-134-000C-5021	Sequence 5021, Ap
43	81.5	4.8	331	1	US-08-134-570-12	Sequence 12, Appl
44	81.5	4.8	479	1	US-08-665-220-2	Sequence 2, Appli
45	81.5	4.8	479	3	US-09-291-692-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-134-001C-4794  
; Sequence 4794, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4794  
; LENGTH: 1037  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-4794

Query Match	6.1%	Score 103.5;	DB 4;	Length 1037;
Best Local Similarity	22.7%	Pred. No. 0.1;		
Matches	61;	Conservative 41;	Mismatches 82;	Indels 85; Gaps 14;
QY	82	EDFRKKTLEQWVDNGFAVLNADNLN-----IATNSOLN--EY-CLSDALQALR	127	
Db	224	KSEIYEYMRDKNDNAIVVNCNMENIDFVGHTGDSIVVAPSQTLSDVEYQMLRDSVSLKVR	283	
QY	128	AYGTGEG-----FEESLWFLDEASKAVKARALQ-----AAMISVDLPFG	168	
Db	284	ALGIEGGCNVOLALPHSLNLYIIEVNPV-SRSSALASKATGYPIAKIAVGLTLD	342	
QY	169	EEFLISP-----ACQNPLKKMVEEVPFR-----APRSTVLYLGDR	206	
Db	343	E-MLNPITGTSYAAFEPLTDYISK-IPRFPDFKFKGERELGTQMKATGEVMAIGRT-	398	
QY	207	GKHSJFPERIEFEVULGTFD-----PHGRMPDL-----ILHDEVRGWLFLMEAVKSG	254	
Db	399	-----YEESLLKAIRSLRYGVHHLGLPNGESVELDYIKERIGHQDDERLFFIGEAIR---	450	
QY	255	PFDEERHESLOELFVTPSAGLIFVNCEN	283	
Db	451	-----RGTSLEELHNMTKIDYFFLNKFN	474	

RESULT 2

US-08-802-141-4  
; Sequence 4, Application US/08802141

```

; Patent No. 5773009
; GENERAL INFORMATION:
; APPLICANT: GLASS, ROGER I.
; APPLICANT: GENTSCH, JOHN R.
; APPLICANT: BHAN, M. K.
; APPLICANT: DAS, BIMAL K.
; TITLE OF INVENTION: ROTAVIRUS STRAIN AND RELATED
; COMPOSITIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,141
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/231,041
; FILING DATE: 15-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.609
; TELEPHONE: 404/688-0770
; TELEFAX: 688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-802-141-4

Query Match 5.9%; Score 100; DB 1; Length 772;
Best Local Similarity 22.6%; Pred. No. 0.15;
Matches 76; Conservative 48; Mismatches 128; Indels 84; Gaps 16;

QY 1 VNSSDIDGTVASIITARALLKRFQDAQRYNVRSAVTLALA-----CLKPGRWVDST 55
DB 473 VPSDDYQTPIANSVTVRODLER-QLDENRREFNELSANIALSOLIDLALLP-----LDMF 527

QY 56 TPRLGVQKIMWSGHEWAKPYATGSRDPRK-----KTLRQ 91
DB 528 SMFSGIRSTIE-----AAKNFATSVMKPKFKSNLAKSVNSLTDAITDAAGSISRSTLRS 582

QY 92 -----WVDNGFAVLNADNINIATNSQLNEYCLSDALQALRAYCTEGF--EESLVFEL 142
DB 583 ANSAVSVMWTDISDVIDSTDNVVTAT-----ATAAAKPRVKKEFTTEFNGVSPD 630

QY 143 DEASKAVKARAEALQAAMTSVDLPGGEEFLLSPAGQNPLLKQWVEFVPRFAPR---STV 199
DB 631 DISAAVKTKMKNLV-----VD-----EEML-----PQITASEKFIENRAYRLIDGDK 676

QY 200 LYLGDTRGKHSIFEREIFEVLGLTFDPHGRMPDLILHDEVGRGLFMEAVKS-KGPFDE 258
DB 677 YVEVTTGKYFYALYLTETEEVM--FDAAE-RFAELVTSYSPVISAIDFKTIKNLDNYGI 732

QY 259 EHRSLQELFVTPSAGLIIFWNCFENRESMRQWLP 294
DB 733 TREQALNMLRSDPKVLRSPIN--QNNPIIKNRIEQL 766

; RESULT 3
;
; APPLICANT: Wetmur, James G.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
; TITLE OF INVENTION: MUTL GENES AND PROTEINS AND USES THEREFOR
; FILE REFERENCE: MSM95-02
; CURRENT APPLICATION NUMBER: US/08/676,444A
; CURRENT FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 511
; TYPE: PRP
; ORGANISM: Thermotoga maritima
; US-08-676-444-42

Query Match 5.7%; Score 96; DB 3; Length 511;
Best Local Similarity 20.8%; Pred. No. 0.21;
Matches 64; Conservative 35; Mismatches 99; Indels 110; Gaps 13;

QY 94 DNGFAVLNADNINIATNSQLNEYCLSDALQALRAYCTEGFEEESLVVFLDEASKAVKARA 153
DB 59 DNGIGMTREAL-LAIEPYTTSKIESEEDLHRTYGFGRGEALASIVQVSRKIVTKTEK 117

US-09-252-991A-21111
; Sequence 21111, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21111
; LENGTH: 310
; TYPE: PRP
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-21111

Query Match 5.8%; Score 97.5; DB 4; Length 310;
Best Local Similarity 24.0%; Pred. No. 0.067;
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;

QY 94 DNGFAVLNADNINIATNSQLNEYCLSDALQALRAYCTEGFEEESLVVFLDEASKAVKARA 153
DB 90 DNKAAYI-----TEVNSQTDFLALQDD---PKGFVAESLEKAFNEKLTDAAPLIVEARE 139

QY 154 EALQAAM-----ISVDLPGEFEFLLSPAGQNPLLK 183
DB 140 EARLALVAKTCNGENVNIRRLTRVEGDVVGYALHGHRIQVVVNLKGG-----NPALA 189

QY 184 KQVEEFV---PRFAPRSTVLYLGDTRGKHSIFEREIFEVLGLTFDPHGRMPDLILHDE 239
DB 190 KDIAMHVAASNPOQLSASEVSEEAIAK-----EKEIF---LALNADKIAGKPNIVENM 240

QY 240 VRGWL--FLMEAVKSKGPFDEERHRSLOELFVTPSAGLIIFVNCFENRESMRQWLP 297
DB 241 VKGRISKFLAEASLVVEQFFVKNPEVKVGDLLAKQAGAEIVSFVRYEVGEGIEKAEVDFAAE 300

QY 298 TEAWVA 303
DB 301 VAAQVA 306

; RESULT 4
;
; APPLICANT: Wetmur, James G.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
; TITLE OF INVENTION: MUTL GENES AND PROTEINS AND USES THEREFOR
; FILE REFERENCE: MSM95-02
; CURRENT APPLICATION NUMBER: US/08/676,444A
; CURRENT FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 511
; TYPE: PRP
; ORGANISM: Thermotoga maritima
; US-08-676-444-42

Query Match 5.7%; Score 96; DB 3; Length 511;
Best Local Similarity 20.8%; Pred. No. 0.21;
Matches 64; Conservative 35; Mismatches 99; Indels 110; Gaps 13;

QY 94 DNGFAVLNADNINIATNSQLNEYCLSDALQALRAYCTEGFEEESLVVFLDEASKAVKARA 153
DB 59 DNGIGMTREAL-LAIEPYTTSKIESEEDLHRTYGFGRGEALASIVQVSRKIVTKTEK 117
```



Qy 102 ADNLNATNSQAEYCL----SDEALQALRAYGTEG-FBESLVVFLDEASKAVKARAEAL 156  
Db 305 AEQATQBVKNWMTETLLVQNSNPDKQILKALGPATLEEMVACQGVGGPTHKAKILAE 364  
Qy 157 QAAMISVDLPGGEBFLLSPAGQNPLLKKMYBEF 189  
Db 365 AMASAOQDLKGGYTAVFMQGNPNRKGPIKCF 397

RESULT 8  
US-08-468-059-59  
; Sequence 59, Application US/08468059  
; Patent No. 5840480  
; GENERAL INFORMATION:  
; APPLICANT: Guertler, Lutz G.  
; APPLICANT: Eberle, Josef  
; APPLICANT: Brunn, Albrecht v.  
; APPLICANT: Knapp, Stefan  
; APPLICANT: Hauser, Hans-Peter  
; TITLE OF INVENTION: Retrovirus from the HIV Group and its  
; TITLE OF INVENTION: Use  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,059  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/132,653  
; FILING DATE: 05-OCT-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 33 646.5  
; FILING DATE: 06-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 35 718.7  
; FILING DATE: 22-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 44 541.8  
; FILING DATE: 30-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 43 18 186.4  
; FILING DATE: 01-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carol P. Einaudi  
; REGISTRATION NUMBER: 32,220  
; REFERENCE/DOCKET NUMBER: 05495-0001-02000  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 498 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-468-059-59

Db 365 AMASAOQDLKGGYTAVFMQGNPNRKGPIKCF 397  
RESULT 7  
US-08-471-770-59  
; Sequence 59, Application US/08471770  
; Patent No. 5770427  
; GENERAL INFORMATION:  
; APPLICANT: Guertler, Lutz G.  
; APPLICANT: Eberle, Josef  
; APPLICANT: Brunn, Albrecht v.  
; APPLICANT: Knapp, Stefan  
; APPLICANT: Hauser, Hans-Peter  
; TITLE OF INVENTION: Retrovirus from the HIV Group and its  
; TITLE OF INVENTION: Use  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,770  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/132,653  
; FILING DATE: 05-OCT-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 33 646.5  
; FILING DATE: 06-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 35 718.7  
; FILING DATE: 22-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 44 541.8  
; FILING DATE: 30-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 43 18 186.4  
; FILING DATE: 01-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carol P. Einaudi  
; REGISTRATION NUMBER: 32,220  
; REFERENCE/DOCKET NUMBER: 05495-0001-03000  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 498 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-471-770-59

Query Match 5.5%; Score 93; DB 1; Length 498;  
Best Local Similarity 26.8%; Pred. No. 0.42;  
Matches 41; Conservative 20; Mismatches 68; Indels 24; Gaps 6;  
Qy 48 GD---RWVDSTTFRGLGVQKIMDSGEHWAKPVATGSRDFFR---KKTLRQWVDNGFAVLN 101  
Db 258 GDIIYRKWI-----VLGLNKNW-----RWVSFVSIIDIROGPKPEFRDYVDRFYKTLR 304

Query Match 5.5%; Score 93; DB 2; Length 498;  
Best Local Similarity 26.8%; Pred. No. 0.42;

	Matches	41;	Conservative	20;	Mismatches	68;	Indels	24;	Gaps	6;
Qy	48	GD----	RWDDSTPRLGVQKIMDWSGEHWAKPYATGSRDEFR---	KKTILROWVDNGFAVLN	101	:	:	:	:	:
Db	258	GDIYRKWI-----	VLGLENKWV-----	KMYSPTSILDTIQGPKPEPRDYVFYKTLR	304	:	:	:	:	:
Qy	102	ADNLNIATNSQLNBVCYL-----	SDEALQALRAYGTG--FEESLVVFLDEASKAHVRAAEAL	156	:	:	:	:	:	:
Db	305	AEQATQEVKNWMTETLLVQNSNPCKQLIKALGFPEATLEENMWACQGVGGFTHKAKILAE	364	:	:	:	:	:	:	:
Qy	157	QAAMISVDLPGGBEEFLSPAGQNPLLKXWVEEF	189	:	:	:	:	:	:	:
Db	365	AMASAQQDLXGGYTAVFMQRGNPNRKGPICKCF	397	:	:	:	:	:	:	:

RESULT 9  
US-09-109-916-59  
; Sequence 59, Application US/09109916

```

; Patent No. 6277561
;
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brunn, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
;
; TITLE OF INVENTION: RETROVIRUS FROM HIV
;
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/109,9
; CURRENT FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: DE P 42 33
; EARLIER FILING DATE: 1992-10-06
; EARLIER APPLICATION NUMBER: DE P 42 35
; EARLIER FILING DATE: 1992-10-22
; EARLIER APPLICATION NUMBER: DE P 42 44
; EARLIER FILING DATE: 1992-12-30
; EARLIER APPLICATION NUMBER: DE P 43 18
; EARLIER FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRN
; ORGANISM: Human immunodeficiency virus
; US-09-109-916-59

```

Query Match 5.5%; Score 93; DB 3; Length 498;

Matches	41; Conservative	20; Mismatches	68; Indels	24; Gaps	6;	
QY	48	GD	---RWVDSTT	PLRLGVKIMDSGEHWAKPYATGSRDEPR	---KKTLRQWDVNDGFAVLN	101
			::: :::	::: :::	::: :::	
Db	258	GDIYRKWI	-----VLGLNKKV	-----KMYSPVSILDIRQGPKEPRFYDVFYKTLR		304
QY	102	ADNLNTATNSQLNEYCL	-----SDEALQALRAYGTGEG	-FEESLVVFLDEASKVAKARAEAL		156
		::: :::	::: :::	::: :::	::: :::	
Db	305	AEQATQEVKNWMTETLLVQNSNPDCQILKALGPEATLEEMWACQGVGPTHKAKILAE				364
QY	157	QAAMISVDLPGGSEFLLSPAGQNPILLKQWVEEF				189
		::: :::	::: :::	::: :::	::: :::	
Db	365	AMASAOODLKGGYTAVFMORGQNPNRKGPICKCF				397

RESULT 10  
US-09-886-156-59  
; Sequence 59, Application US/09886156  
; Patent No. 6528626  
; GENERAL INFORMATION:  
; APPLICANT: Guertler, Lutz G.  
; APPLICANT: Eberle, Josef  
; APPLICANT: Brunn, Albrecht V.  
; APPLICANT: Knapp, Stefan  
; APPLICANT: Hauser, Hans-Peter  
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND

```

; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,156
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US/09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-886-156-59

```

Query Match 5.5%; Score 93; DB 4; Length 498;  
Best Local Similarity 26.8%; Pred. No. 0.42;  
Matches 41; Conservative 20; Mismatches 68; Indels

Qy	48	GD---	RWVDS	TPTPLG	VQKIMD	WSGEHWAKPYATGS	REDFR---	KRTLRQWDNGEFAVLN	101
		:	:	:	:	:	:	:	:
Dd	258	GDIYRKWI	-----	VIGLNKW	-----	KMVSPTSILDIRQPKPEFFDYDRFVKTLR			304
		:	:	:	:	:	:	:	:
Qy	102	ADNLNIATNSQLNEYCL	----	SDEALQALRAYGTGEG	-FEESIVFLDEASKAVKARAEL				156
		:	:	:	:	:	:	:	:
Dd	305	AEOATQEVENMTETLLVQNSNPDCQIKLKGALGPATLEEMMVACQGVGPTHAKILAE							364
		:	:	:	:	:	:	:	:
Qy	157	QAAMISVDLPGGREFFLLSPAQCNPLLKKQWEEF							189
		:	:	:	:	:	:	:	:
Dd	365	AMASAOQDLKGGYTAVFMORCONPNRKGPICPF							397
		:	:	:	:	:	:	:	:

## RESULT 11

```

US-09-886-149-59
; Sequence 59, Application US/09886149
; Patent No. 6531137
; GENERAL INFORMATION:
; APPLICANT: Guertler, Lutz G.
; APPLICANT: Eberle, Josef
; APPLICANT: Brumm, Albrecht V.
; APPLICANT: Knapp, Stefan
; APPLICANT: Hauser, Hans-Peter
; TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
; FILE REFERENCE: 05495.0001-04
; CURRENT APPLICATION NUMBER: US/09/886,149
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/109,916
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: DE P 42 33 646.5
; PRIOR FILING DATE: 1992-10-06
; PRIOR APPLICATION NUMBER: DE P 42 35 718.7
; PRIOR FILING DATE: 1992-10-22
; PRIOR APPLICATION NUMBER: DE P 42 44 541.8
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: DE P 43 18 186.4
; PRIOR FILING DATE: 1993-06-01
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus
US-09-886-149-59

```

Query Match 5.5%; Score 93; DB 4; Length 498;  
Best Local Similarity 26.8%; Pred. No. 0.42;  
Matches 41; Conservative 20; Mismatches 68; Indels



Search completed: October 1, 2004, 16:13:58  
Job time : 15.1755 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:13:07 ; Search time 69.0717 Seconds  
(without alignments)  
1504.830 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSDDIGTGVASIDTARAL.....EDPDHLHLNGSRFLGPYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues  
Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	7.0	708	12	US-10-627-476-314
2	118	7.0	713	12	US-10-627-476-312
3	118	7.0	1111	9	US-09-738-626-6581
4	105.5	6.2	229	15	US-10-211-689-34
5	105.5	6.2	236	15	US-10-211-689-42
6	103.5	6.1	1120	12	US-10-282-122A-70903
7	103	6.1	194	15	US-10-211-689-30
8	101	6.0	230	15	US-10-211-689-46
9	99.5	5.9	217	15	US-10-211-689-44
10	99.5	5.9	1057	12	US-10-282-122A-71880
11	98.5	5.8	782	16	US-10-437-963-148885
12	98.5	5.8	1493	16	US-10-437-963-132386
13	97.5	5.8	289	9	US-09-815-242-11940
14	97.5	5.8	289	12	US-10-282-122A-66521
15	97	5.7	485	15	US-10-369-493-17102

16	96	5.7	516	15	US-10-369-493-2886	Sequence 2886, Ap
17	95.5	5.7	797	9	US-09-815-242-5071	Sequence 5071, Ap
18	95.5	5.7	797	12	US-10-282-122A-43545	Sequence 43545, A
19	95.5	5.7	1057	12	US-10-282-122A-70305	Sequence 70305, A
20	94	5.6	155	9	US-09-738-626-5979	Sequence 5979, Ap
21	93.5	5.5	454	15	US-10-369-493-8439	Sequence 8439, Ap
22	93	5.5	498	9	US-09-886-156-59	Sequence 59, Appl
23	93	5.5	498	9	US-09-886-150-59	Sequence 59, Appl
24	93	5.5	498	10	US-09-886-149-59	Sequence 59, Appl
25	93	5.5	498	10	US-09-886-159-59	Sequence 59, Appl
26	93	5.5	498	14	US-10-326-090-59	Sequence 59, Appl
27	92.5	5.5	233	15	US-10-211-689-32	Sequence 32, Appl
28	92.5	5.5	239	15	US-10-211-689-36	Sequence 36, Appl
29	92	5.4	445	9	US-09-815-242-5010	Sequence 5010, Ap
30	92	5.4	447	9	US-09-815-242-10908	Sequence 10908, A
31	92	5.4	447	12	US-10-282-122A-42511	Sequence 42511, A
32	91.5	5.4	444	12	US-10-282-122A-45519	Sequence 45519, A
33	91.5	5.4	685	16	US-10-437-963-118346	Sequence 118346
34	91.5	5.4	1222	16	US-10-408-765A-1650	Sequence 1650, Ap
35	91.5	5.4	1360	9	US-09-788-657-22	Sequence 22, Appl
36	91.5	5.4	1360	10	US-09-912-697-16	Sequence 16, Appl
37	91.5	5.4	1360	10	US-09-760-285-26	Sequence 26, Appl
38	91.5	5.4	1360	14	US-10-270-839-37	Sequence 37, Appl
39	91.5	5.4	1360	14	US-10-371-634-15	Sequence 15, Appl
40	91.5	5.4	1360	14	US-10-348-074-40	Sequence 40, Appl
41	91.5	5.4	1360	16	US-10-641-068-22	Sequence 22, Appl
42	91.5	5.4	1360	16	US-10-408-765A-226	Sequence 226, Appl
43	90.5	5.4	503	13	US-10-027-806-66	Sequence 66, Appl
44	90.5	5.4	503	13	US-10-034-623-66	Sequence 66, Appl
45	90.5	5.4	503	14	US-10-027-801-66	Sequence 66, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-627-476-314  
; Sequence 314, Application US/10627476  
; Publication No. US20040030116A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Mark  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schoder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Habernauer, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
; TITLE OF INVENTION: TRANSPORT  
; FILE REFERENCE: BGI-125CPCN  
; CURRENT APPLICATION NUMBER: US/10/627,476  
; PRIOR FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: 09/602,787  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: USSN 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: DE 19931454.3  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931478.0  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931563.9  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19932122.1  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932124.8  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932125.6  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932128.0  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932180.9  
; PRIOR FILING DATE: 1999-07-09  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 678

```
; SEQ ID NO 314
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-627-476-314

Query Match          7.0%; Score 118; DB 12; Length 708;
Best Local Similarity 25.5%; Pred. No. 0.0093;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLLALAGLPGDRWVDSTTPRLGVQKIMDSGE---HWAK-----P 75
Db 172 KRFNRCVPTPEELSALGPRDGVTI----LGVRFQAQQTGPTTRWEKDRRKLGDST 227
QY 76 YATGSRDFRKKTLROWVDNGFAVLN-ADNLNIATNSQLNE---YCLSDALQALRAYG 130
Db 228 YRLGSTNDKAVETTLRETQKAGVAVQAADNRRAANRAELRELRQYQASQELKVSQAQI 287
QY 131 TEGFEESLVVFLD-----EASKAVKARAEALQAAAMISVDLPGGGEF 171
Db 288 DVESADAAIAELDRLLBELNNTPEATLSARHEAAKQTLARVSDLLVAAQSEETVASMNL 347
QY 172 LLSPAGQNPLKKMWEEFVPFAPRSTVLYLGDTRGKHSLEFEEVVLGTFD 226
Db 348 KRAETELKRLSLPVAEVSEIEAREVEKFLANTRRVHA---ANVDEQITIALRED 399

RESULT 3
US-09-738-626-6581
; Sequence 6581, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKADA, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6581
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6581

Query Match          7.0%; Score 118; DB 9; Length 1111;
Best Local Similarity 25.5%; Pred. No. 0.018;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLLALAGLPGDRWVDSTTPRLGVQKIMDSGE---HWAK-----P 75
Db 570 KRFNRCVPTPEELSALGPRDGVTI----LGVRFQAQQTGPTTRWEKDRRKLGDST 625
QY 76 YATGSRDFRKKTLROWVDNGFAVLN-ADNLNIATNSQLNE---YCLSDALQALRAYG 130
Db 626 YRLGSTNDKAVETTLRETQKAGVAVQAADNRRAANRAELRELRQYQASQELKVSQAQI 685
QY 131 TEGFEESLVVFLD-----EASKAVKARAEALQAAAMISVDLPGGGEF 171
Db 686 DVESADAAIAELDRLLBELNNTPEATLSARHEAAKQTLARVSDLLVAAQSEETVASMNL 745
QY 172 LLSPAGQNPLKKMWEEFVPFAPRSTVLYLGDTRGKHSLEFEEVVLGTFD 226
Db 746 KRAETELKRLSLPVAEVSEIEAREVEKFLANTRRVHA---ANVDEQITIALRED 797

; SEQ ID NO 312
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-627-476-312

Query Match          7.0%; Score 118; DB 12; Length 708;
Best Local Similarity 25.5%; Pred. No. 0.0093;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLLALAGLPGDRWVDSTTPRLGVQKIMDSGE---HWAK-----P 75
Db 167 KRFNRCVPTPEELSALGPRDGVTI----LGVRFQAQQTGPTTRWEKDRRKLGDST 222
QY 76 YATGSRDFRKKTLROWVDNGFAVLN-ADNLNIATNSQLNE---YCLSDALQALRAYG 130
Db 223 YRLGSTNDKAVETTLRETQKAGVAVQAADNRRAANRAELRELRQYQASQELKVSQAQI 282
QY 131 TEGFEESLVVFLD-----EASKAVKARAEALQAAAMISVDLPGGGEF 171
Db 283 DVESADAAIAELDRLLBELNNTPEATLSARHEAAKQTLARVSDLLVAAQSEETVASMNL 342
QY 172 LLSPAGQNPLKKMWEEFVPFAPRSTVLYLGDTRGKHSLEFEEVVLGTFD 226
Db 343 KRAETELKRLSLPVAEVSEIEAREVEKFLANTRRVHA---ANVDEQITIALRED 394

RESULT 2
US-10-627-476-312
; Sequence 312, Application US/10627476
; Publication No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-125CPCN
; CURRENT APPLICATION NUMBER: US/10/627,476
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: 09/602,787
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: USSN 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 312
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-627-476-312

Query Match          7.0%; Score 118; DB 12; Length 713;
Best Local Similarity 25.5%; Pred. No. 0.0094;
```

```
RESULT 4
US-10-211-689-34
; Sequence 34, Application US/10211689
; Publication No. US20030232347A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khrastov, Nikolai V.
; APPLICANT: Lepley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-416B
; CURRENT APPLICATION NUMBER: US/10/211,689
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 34
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-34

Query Match      6.2%; Score 105.5; DB 15; Length 229;
Best Local Similarity 23.1%; Pred. No. 0.034;
Matches 56; Conservative 39; Mismatches 114; Indels 33; Gaps 11;

QY 35 SAVTLLAAGLPGRWVDSTTRGLGVQKIMDW-SGEHWAKPYATGSRDFRKTLRQWV 93
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6 AALLVTFLAGCQAKVEQVETPEPELROQTQWQSQRW--ELALGRFWDYLR-----WV 58
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 94 DNGFAVLNADNLNIATNSOLNEYCLSDALQALRAYGTGFPESLVVFLDEASKAVKAPA 153
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
59 QTLRSQVQELUSSQVTOELR--ALMDETMKELKAYKSE--LEEQLTPVAETPRLSKEL 115
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 154 EALQAMISVDLPGGGEFFLLSPAGQ-NPLLKKWVEFVPRFAP-----RSTVLVLGD--- 204
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 116 QAAQ-AELGADMEDVRGRLVQYRGEVQAMLGQSTEEELRVRLASHLRKURKLLRDADLE 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 TRGHKSIFEREIFEVELGLTFDPHGRMPDLILHDEVRGWLFLMEAVK-----SKGPFDEE 259
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 EQAQIIRLQAEAFQARLKSWFEP-----LVEDMQRWAGLVKVKVQAAGVTSAAAPVPSD 227
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 RH 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 NH 229

RESULT 5
US-10-211-689-42
; Sequence 42, Application US/10211689
; Publication No. US20030232347A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khrastov, Nikolai V.
; APPLICANT: Lepley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-416B
; CURRENT APPLICATION NUMBER: US/10/211,689
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 42
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-42

Query Match      6.2%; Score 105.5; DB 15; Length 236;
```

[illegible]

## RESULT 6

```

US-10-282-122A-70903
; Sequence 70903, Application US/10282122A
; Publication NO. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.0344
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70903
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70903

```

[illegible]

## RESULT 7

```

US-10-211-689-30
; Sequence 30, Application US/10211689
; Publication No. US2003023347A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khamatov, Nikolai V.
; APPLICANT: Lepley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTI
; FILE REFERENCE: 21402-416B
; CURRENT APPLICATION NUMBER: US/10/211,689
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13

```

```

; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data remain
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 30
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-30

```

Query Match 6.1%; Score 103; DB 15; Length 194;  
Best Local Similarity 21.9%; Pred. No. 0.048;  
Matches 51; Conservative 37; Mismatches 95; Indels 50; Gaps 10;

[illegible]

QY 94 DNGFAVLNADNLNIATNSQLNECYCLSDLEALQALRAYGTEGFEESLVFLDEASKAVKARA 153

Db 59 QTLSEQVEELLSSQVTQELR--ALMDETMKELKAYKSE-LEEQLTPVAEETRARLSKEL 115

QY 154 EALQAAMISVDLPGGEEFLSPAGQNPLLKKMVEEFVPRFAPRSTVLYLGDTRGKHSLEE 213

[illegible]

22 110 QAAQAAU-----GADEV-----NGQVAEVRAKLEE-----QAQQIRLQ 148

QY 214 REIFEEVLGTFDPHGRMPDLILHDEVRGWFLMEAVK-----SKGPFDEERH 261

Db 149 AEAFAQRLKSWFEP-----LVEDMQROWAGLVEKVOAAVGTSAAPVPSDNH 194

## RESULT 8

US-10-211-689-46

Sequence 46, Application US/10211689  
Publication No. US20030232347A1  
GENERAL INFORMATION:  
APPLICANT: Alsbrook, John II  
APPLICANT: Anderson, David W.  
APPLICANT: Boldog, Ferenc L.  
APPLICANT: Burgess, Catherine E.  
APPLICANT: Casman, Stacie J.  
APPLICANT: Edinger, Shlomit R.  
APPLICANT: Gangolli, Esha A.  
APPLICANT: Gorman, Linda  
APPLICANT: Guo, Xiaojia (Sasha)  
APPLICANT: Khramstov, Nikolai V.  
APPLICANT: Lepley, Denise M.  
APPLICANT: MacDougall, John R.  
APPLICANT: Pena, Carol A.  
APPLICANT: Peyman, John A.  
APPLICANT: Patturajan, Meera  
APPLICANT: Rieger, Daniel K.  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Smithson, Glennda  
APPLICANT: Spytek, Kimberly A.  
APPLICANT: Vernet, Corine A. M.  
APPLICANT: Voss, Edward Z.  
APPLICANT: Zhong, Mei  
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
FILE REFERENCE: 21402-416B  
CURRENT APPLICATION NUMBER: US/10/211,689  
CURRENT FILING DATE: 2003-01-21  
PRIOR APPLICATION NUMBER: 60/311751  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: 60/310,802  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: 60/310,795  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: 60/311,292  
PRIOR FILING DATE: 2001-08-09

```

; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: Curaseq1 version 0.1

```

```

; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-46

```

Query Match 6.0%; Score 101; DB 15; Length 230;  
Best Local Similarity 23.2%; Pred. No. 0.1;  
Matches 53; Conservative 39; Mismatches 108; Indels 2

35 3A1111 Y AT ACT VBCDEFGHCEETEDY CHOCYSTYMDA

02 33 SAVIDUAGLNFGRWVDSIERGLGVQKIMDW-SGEHWAKFYATGSRDEFKKTILRQWV 93

Db 8 AALLVTFLAGCQAKVEQAVETEPELRRQQTWQSGQRW--ELALGRFWDYLR-----WV 60

QY 94 DNGFAVLNADNLNIATNSQLNEYCLSDLEALQALRAYGTGFEESLVVFLDEASKAVKARA 153

21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 54

## Qy

Db 118 QAAQ-ARLGADMEDVGRGLVQYRGVEQAMLGQSTEEIYVFLASHLKKLRKLLRDADDLE 176

QY 205 TRGKHSLEFEREIFEVLGLTFDPGRMPDLLIHDEVRGWLFLMEAVKS 252

Db 177 EQAQOIRLQAEAFARLKSWFEP-----LVEDMORWAGLVKVKVOA 217

## RESULT 9

```

US-10-211-689-4
; Sequence 44, Application US/10211689
; Publication NO. US2003023247A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gangolli, Bsha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khramstov, Nikolai V.
; APPLICANT: Lephley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Smithson, Glenda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-4168
; CURRENT APPLICATION NUMBER: US/10/211.689
; CURRENT FILING DATE: 2003-01-21

```

```

; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 44
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-211-689-44

```

```

Query Match      5.9%; Score 99.5; DB 15; Length 217;
Best Local Similarity 23.4%; Pred. No. 0.13;
Matches 52; Conservative 35; Mismatches 102; Indels 33; Gaps 11;

Qy 55 TTPRLGVQKIMDW-SGEHWAKPYATGSRDPFRKTLROWVNGFAVLNADNINATNSQL 113
Db 11 TEPEPRLQQTQWQSGQRW--ELALGREFWDYLR-----WVQTLSEQVQEEILLSQVTOEL 63

Qy 114 NEYCLSEALQALRAYGTGFEFSIVVFLDEASKAVKARAEALQAMTISVLPQGEFLL 173
Db 64 R-ALMDTEMLKELKAYKE-JEEQTPVAETRAKLSKELQAAQ-ARLGADMEDVGRGLV 119

Qy 174 SPAGQ-NPLKKVMVEEFVRFPAP-----RSTVLYLGD---TRGKHSLEFERIFEVLGLT 224
Db 120 QYRGEVQALQSGTEELRVRLASHLRLKRLRDADDLEBQAOQIRLQAEAFQARLKS 179

Qy 225 FDPGRMPDLIIHDEVGRWLFMEAVK-----SKGPFDEERH 261
Db 180 FEP-----IVEDMQRWAGLVEKVQAAVGTSAAPVPSDNH 214

```

```

RESULT 10
US-10-282-122A-71880
; Sequence 71880, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71880
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
; US-10-282-122A-71880

```

```

Query Match      5.9%; Score 99.5; DB 12; Length 1057;
Best Local Similarity 21.9%; Pred. No. 1.4;
Matches 59; Conservative 42; Mismatches 83; Indels 85; Gaps 14;

Qy 82 EDRFKTLROWVNGFAVLNADNIN-----IATNSQLN-EY-CLSDALQALR 127
Db 214 KEIYEVRMDKNDNAIVVCNMENIDPVGIHTGDSIVVAPSQTLSDEVQMLRDLVKVIR 273

Qy 128 AYGTGEG-----FEESLVFLDEASKAVKARAEALQ-----AAMISVDLPFG 168
Db 274 ALGTGEGCNVOLADPHSFNYIIIEVNPVRV-SRSALASKATGPIAKLAKIAVGLTLD 332

Qy 169 EEFLLSP-----AGQNPLLKKVMVEEFVRPF-----APRSTVLYLGDTR 206
Db 333 E-MLNPTVGTSYAAFEPTLDYVISK-IPRPFDFKFKGERELGTQMKATGEVMAIGRT- 388

Qy 207 GKHSLEFERIFEVLGLTDFD-----PHGRMPDL-----LLHDEVGRWLFMEAVKSKG 254
Db 389 -----YEESLLKAIKRSLEYGVHGLPNGESFDLDYIKERISHQDDERLFFIGEAIR--- 440

Qy 255 PFDERHRSLOELFVTPSPAGLIFVNCFFEN 283
Db 441 -----RGTTLEIHNMTOIDYFFLNKFN 464

```

```

RESULT 11
US-10-437-963-148885
; Sequence 148885, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 148885
; LENGTH: 782

```

```

; TYPE: PR1T
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_49272C.1.pcp
US-10-437-963-148885

```

Query Match 5.8%; Score 98.5; DB 16; Length 782;  
Best Local Similarity 22.2%; Pred. No. 1.1;  
Matches 83; Conservative 37; Mismatches 125; Indels 129; Gaps 20;

Qy	15	DTARALLKBEFGDQRYNVRSAVTLALAGLKPGDRWVDSTTPRLGVQKIMDSG-EHWA	73
Db	142	DTA-----VRXFKDVMMMIRFVAMVLDTLGEASIDRN-----RLGTSRLNIGGFGEHAF	195

Qy	74	KPATGSGREDFRK-	-----TLQWVD-----	NGFA	98
				:	
Db	196	YKYPL-----	NFRKNSNPNDPSATTFVTITWRDQPEAGSDIAFVLSSTNKLINHS	251	

Qy	99	-----VLNADN-----		-----LN-----	-----IATNSQLNEYCLSD	120
		:				
Db	252	LGQYLGLEFNASNTSQNTLAI	ELDTFM	PNLMDMDNDHVG	IVDVALSINSHTAGFTYSD	311

Cy            121   EALQALRAYGTEGEESLVVFLDEASKVKARAEALQAAMISVDLPGGEEFLSPAGNP   180  
Db            312   GGFQLRL--ANGRSPILQLWVDYDKAHQLN-----VTGLP----YSKPEYP   355

QY		181	LLKKMBEEFVPAPRSTVILYGDTGRGKHSLFEREIFEVLGLTFDPHGMRPDL----	IL 236
			:::    :	
Dd		356	LSSIVN--LSSLPPSS--YIGFSASVNSPKTRHF---ILGWSEKENGVRPLPSVPVT	408
			:::    :	

QY 237 HDEVRCW---LFLMEAYVSKGPPDEERHRSLOEL--FVTPSAGLIFVNCFNRESMRQWL 291

Dd 409 DPETYGGGNFFAPPPPPQLNTHQVHKH-SLQILPIVMTSVILLLVAF----- 457

Qy	292	PELAWETEANVAED	305
Db	458	--LGWRKKAGPOED	469

```

RESULT 12
US-10-437-963-132386
; Sequence 132386, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21153221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 132386

```

```

/
/ ORGANISM: Oryza sativa
/
/ FEATURE:
/
/ NAME/KEY: unsure
/
/ LOCATION: (1)..(1493)
/
/ OTHER INFORMATION: unsure at all Xaa locations
/
/ FEATURE:
/
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_3435C.1.pep
US-10-437-963-132386

```

Query Match	5.8%	Score 98.5;	DB 16;	Length 1493;
Best Local Similarity	21.0%;	Pred. No. 3;		
Matches	78;	Conservative	41;	Mismatches 126;
				Indels 127;
				Gaps 17;

QY	3	SSDGIDGTVASID-----TARALKRFGFDAQRYNVRSVATLLALAGLKPGRW-----	51
Db	542	ADDGAQGTVATLDGGGGXCAHTSRRCFKAMRRELLAAVPTHEAARKA--RWSEVKLT	598
QY	52	-----VDSTTPRLGVQKIMDSEHNAKPYATGSRDEFKTKTLRQWVDNGFA	98
Db	599	FQDSDHPTAPGMKLOPSLPIIGVTGHTWPLGHVELPMTFGDSTNFRTER-----	652
QY	99	VLNADNLNIATNSQINCYCL-----SDEALQALRAYGTEGFEESLVVFLD-----	143
Db	653	VA---DLNLPCYNVLGRPTLVKFMAATHYAVLQMKMQGPAG--PITVFGDVKVALTCAE	706
QY	144	-----EASKAVKARAEA-LQAAMISVDLPGCEBFLSLPAGQNPLLKXM	185
Db	707	QCADNLAVATEPQAPASASRXXKKRLTSADDEGALVSFLRANSDFAMKLSMPRVIXEV	766
QY	186	VEEVVPFAPRSTVLYLGDTRGKSLPERETFEVLGLTFDPDHGRMPDLILHDEVGRMI-	244
Db	767	IEH---RFAVRQKV-----RQALERQAF-----IREEV-AMLL	796
QY	245	---FLMEAVKSGKGFDEERHRSLOELFTVTPAGLIFVNCFNENRSMRWQLPELAWETEA	301
Db	797	EADFIREVHPW-----LANLVVAPXANGKLRMCIDYTD-LNKACP-----	837
QY	302	VAEDPDHLIHLN	313
Db	838	--KDPFPLPHID	847

RESULT 13

```

US-09-815-242-11940
; Sequence 11940, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11940
LENGTH: 289
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11940

```

Query Match 5.8%; Score 97.5; DB 9; Length 289;  
Best Local Similarity 24.0%; Pred. No. 0.32;  
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;





Mon Oct 4 09:05:16 2004

us-10-668-047-4.rapb

Page 9

Search completed: October 1, 2004, 16:28:47  
Job time : 71.0717 secs

Blank